

Optimization Model Building in Economics

By

Richard E. Howitt

January 2002

ARE 252
Department of Agricultural Economics
University of California, Davis

VISIT...

LANZAROTE
Caliente.COM

<u>Contents</u>		Pages
I.	An Introduction to Optimization Modeling	3-8
II	Specifying Linear Models	9-25
III	Solving Linear Models	26-40
IV	The Dual Problem	41-55
V	Calibrating Optimization Models	56-77
VI	Using Nonlinear Models for Policy Analysis	78-92
VI	Incorporating Risk and Uncertainty	93-101
VIII	An Introduction to Maximum Entropy Estimation	102-116
IX	Nonlinear Optimization Methods	117-137

I An INTRODUCTION to LINEAR MODELS

Definition of a Model

Everyone uses models to think about complex problems. Usually our model is a simple weighting of past experiences that simplifies decisions. For example, after an initial learning period most people drive a car with a model that assumes a certain steering and braking action and only make radical changes from an established pattern when an unexpected emergency occurs. After the emergency most drivers return to their basic model. Why is the model of driving by exception normally optimal? The answer is that this driving model reduces the number of standard decisions that we have to think about and allows us to be more observant for the exceptional situation that requires a different action.

It is often thought that models are limited to algebraic representations and, as such, are hard to construct or interpret. This puts up an artificial barrier to mathematical models that often prevents an evolutionary approach to thinking about them. In reality, everyone uses models to think about complex events, as the process of constructing a model is part of the human process of thinking by analogy. For example, many people use astrology to guide their decisions, a curious but ancient model of relating the position of planets and stars to events in their lives. Skeptics point out that the ambiguity of most astrological forecasts makes quantitative measures hard to confirm. Perhaps they miss the point of astrology, which may not be to accurately predict events, but to give illusion of knowledge over unpredictable events. However, as economists we should be interested in astrology as a product for which the demand has been strong for several millennia.

For this course, the point is to see mathematical models as a practical extension of the graphical models with which we started our micro economic analysis. Mathematical models allow us to explore many more dimensions and interactions than graphical representations, but often we can usefully use simple graphical examples to clarify a mathematical problem. With their larger number of variables, mathematical models can be specified in a more realistic manner than graphical analysis but are still limited by data and computational requirements.

A model is by definition abstracted and simplified from reality and should be judged by its ability to deliver the required precision of information for the task in hand. It is easy in economics to judge a model on its mathematical elegance or originality, that is, as a work of art or artifact rather than a tool

Types of Models

Verbal Models

Thomas Kuhn has proposed that most scientific thought takes place within paradigms that gradually evolve. Given the evolutionary nature of science it is not surprising that most research takes place within a paradigm rather than trying to change paradigms. One of the older and best-known paradigms in economics is Smith's analogy of the price and market system to an "invisible hand". Simple verbal models such as this are very helpful in concisely defining the qualitative properties of a paradigm. The ability to give a simple verbal explanation of the model is probably a necessary condition for full understanding of a complex mathematical model. If

you are unable to explain the essence of what you are modeling to your Grandmother, you probably don't really understand it.

Geometric (Graphical) Models

Geometric methods are the way that we are introduced to economic models and a method where our natural spatial instincts are more easily harnessed to show interrelationships between functions and equilibria. For most empirical applications, graphical models in two or three dimensions are simply too small to adequately represent the empirical relationships needed. Most graphical models can be represented by a system of two equations, whereas optimization models that we will encounter later in the course can have tens or hundreds of equations and variables. However, like verbal models, graphical models are very useful for conceptualizing mathematical relationships before extending them to multidimensional cases.

Algebraic Models

In economics, the term model has become synonymous with algebraic models since they have been the essential tools for empirical and theoretical work for the past five decades. For this course, a critical difference in model specification is between **optimizing behavioral models** and **optimizing structural models**.

Behavioral models yield equations that describe the outcome of optimizing behavior by the economic agent. Assuming that optimizing behavior of some sort has driven the observed actions allows us to deduce values of the parameters that underlie it. For example, observations on the different amounts of a commodity purchased as its price changes, allows the specification and estimation of the elasticity of demand.

An alternative approach to specification of this problem is to define structural equations for the consumer's utility function, represent the budget constraint and alternative products as a set of constraint equations, and explicitly solve the resulting utility optimization problem for the optimal purchase quantity under different prices. With a deterministic model and a full set of parameters, both these approaches would yield the same equilibrium. However this situation rarely occurs and each approach has its relative advantages.

Another distinction is between **positive and normative models**. Behavioral models are invariably positive models where the purpose is to model what economic agents actually do. In contrast, structural models are often normative and are designed to yield results that show what the optimal economic outcome should be. Inevitably normative models require some objective function that purports to represent a social objective function. This is very difficult to specify without strong value judgments.

The Development of Computational Economics

In the past it has usually been the case that econometric models are positive and programming models are normative as they have an explicitly optimized objective function. This has led to an unfortunate methodological division among empirical modelers on the supply side of Agricultural and Development economics into practitioners of econometric and programming approaches. The difference in approach has been divided along the lines of normative and positive models in the past. With the development of calibration techniques for optimization models, programming approaches can now incorporate some positive data based parameters and thus build a continuous connection from the pure data based econometrically estimated models through to the linearly constrained programming models. From one viewpoint econometric models are data intensive while calibrated nonlinear optimization models are more computationally intensive.

In recent years the sub-discipline of Computational Economics has emerged, principally in the empirical application of macro-economic models. The leading text in the area is

Judd(1999) "Numerical Methods in Economics"

As we would expect, shifts in both the supply and demand have stimulated the emergence of this new economic field. The shift in the supply function of computation ability and cost has largely been driven by "Moore's Law" which states that "The number of transistors on a given chip doubles every eighteen months without any increase in cost". This remarkable trend which was first proposed by Gordon Moore, a cofounder of Intel, is predicted to continue at least until the next century. Clearly we are in the middle of a dramatic reduction in the cost of computation. Along with the changes in hardware supply, there have been similar changes stimulated in the supply of software for computational economics.

The demand for computational economics is also shifting out due to the increasing complexity and speed required for applied economic analysis. In addition, several of the newer methodologies in stochastic dynamics and game theory are not suited to the analytical process of deriving testable hypotheses for conventional estimation methods.

Of more concern to those in this course is the fact that growth areas for applied economic analysis are in environmental, resource and development economics. Both these fields are characterized by the absence of large reliable data series and the need for disaggregate analysis. It's not that econometric approaches are unsuited to supply side analysis in these areas, it's just that the data needed is not usually available.

These two shifts bode well for the growth in optimization models in the future. While there are many books on optimization modeling using linear and quadratic structural approaches, for example Paris (1991) or Hazell & Norton (1986). There is no published text on calibrating micro-economic models. This reader is a start on an introductory text for calibrating optimization models.

Sections I - IV are a brief introduction to the specification, solution and interpretation of linear structural models. The specification and solutions are defined in terms of linear algebra for reasons of compactness, clarity and continuity with the remaining sections. Sections V - IX give an introduction to the development of nonlinear calibrated behavioral models.

Uses for Models

Given an economic phenomenon there are three tasks that we may want to perform with economic models. We may wish to explain the observed actions. This is usually performed by

structural analysis using positive econometric models. Given a structure in the form of a specific set of equations, the parameters that most closely explain the observed behavior are accepted as the most likely explanation.

A second practical use for economic models is in predicting economic phenomena. As the Druids found out, forecasting significant economic events is a source of power. Forecasting models are the ultimate outcome of the positivistic viewpoint where the structure is unimportant in itself and the accuracy of the out of sample forecast is the key determinant of model value. Econometric time series models are the best examples of pure forecasting models, although the ability to produce accurate out of sample forecasts should be used to assess the information value of all types of models.

A third use of economic models is to control or influence certain economic outcomes. This process is generally referred to as policy evaluation since public economic policies are justified on the basis of improving some set of economic values. Both structural econometric and optimization models are used for policy evaluation, however due to the dearth of sample data and the wealth of physical structural data; policy models of agricultural production and resource use are often specified as optimization models.

Types of Agricultural Economic Models of Supply

Econometric Models (Positive Degrees of Freedom)

Econometric structural models have been the standard approach to agricultural economic models for the past twenty years. Econometric models of agricultural production offer a more flexible and theoretically consistent specification of the technology than programming models. In addition, econometric methods are able to test the relevance of given constraints and parameters given an adequate data set. The initial econometric research on production models was performed on aggregated data for multioutput / multiinput systems, or single commodities for more disaggregated specifications. However, despite several methodological developments econometric methods are rarely used for disaggregated empirical microeconomic policy models of agricultural production. This is usually because time series data is not generally available on a disaggregated basis and the assumptions needed for cross-section analysis are not usually acceptable to policy makers with regional constituencies. In short, flexible form econometric models have not fulfilled their empirical promise mostly due to data problems that do not appear to be improving.

Constrained Structural Optimization (Programming) Models

Optimization models have a long history of use in agricultural economic production analysis. There is a natural progression from the partial budget farm management analysis that comprised much of the early work in agricultural production to linear programming models based on activity analysis and linear production technology. Often linear specifications of agricultural production are sufficiently close to the actual technology to be an accurate representation. In other cases the linear decision rule implied by many Linear Programming (LP) models is optimal due to Leontief or Von Liebig technology in agriculture.

Despite the emphasis of methodological development for econometric models, programming models are still the dominant method for microanalysis of agricultural production and resource use. Their applications are widespread due to their ability to reproduce detailed

constrained output decisions and their minimal data requirements. As noted above, econometric model applications on a microeconomic basis are hobbled by extensive data requirements.

LP models are also limited largely to normative applications as attempts to calibrate them to actual behavior by adding constraints or risk terms have not been very successful.

Calibrated Positive Programming Models (Zero Degrees of Freedom)

Much of this course is focused around a method of calibrating programming models in a positive manner that has been a major focus of my research over the past ten years (Howitt 1995). The approach uses the observed allocations of crops and livestock to derive nonlinear cost functions that calibrate the model without adding unrealistic constraints. The approach is called Positive Mathematical Programming (PMP). The focus of the course is on specifying, solving and interpreting several Positive and Normative Programming models used in Agricultural and Environmental Economics

Computable General Equilibrium (CGE) Models

CGE models have been used in macro-economic and sectoral applications for the past fifteen years, using a combination of fixed linear proportions from Social Accounting Matrices (SAMS) and calibrated parameters from exogenous elasticities of supply and demand. CGE models have much in common with PMP models in their data requirements and conceptual calibration approach. They will not be addressed directly in this course due to time constraints. Those interested in developing these skills should take ARE 215C

Ill-Posed Maximum Entropy Models (Negative Degrees of Freedom)

This class of models is newly emerging and beyond the scope of this course. Briefly this approach enables consistent reconstruction of detailed flexible form models of cost or production functions on a disaggregated basis. This requires that the model contain more parameters than there are observations – hence the term "Ill-Posed" problems. An application to micro production in agriculture is found in Paris & Howitt AJAE February 1998.

Criteria for Model Selection

Selection of the best model for the research task at hand is an art rather than a science. The model builder is constantly balancing the requirements of realism that complicate the model specification and solution against the practicality of the model in terms of its data and computational requirements. This trade-off is similar to the selection of the optimal photographic models for a mail order catalog where the publisher has to make the subjective trade off between the beauty of the model and the degree of realism that the model will portray. The optimal customer response to the catalog will come from models who are eye-catching but with whom the customers can identify. In economic policy models, they must be simple enough so that the decision maker can identify with the model concept, but at the same time be tractable and able to reproduce the base year data.

There is no ideal model, just some that are more manageable and useful than others. Hopefully, this course will give you the theoretical and empirical tools to make informed decisions on the best model specification for particular data and research situations.

The process of econometric model building has three well-defined stages: Specification, Estimation, and Simulation. Programming model building methods have not formally separated these stages. The equivalent stages are Specification, Calibration and Policy Optimization. However the important process of calibrating the models is usually buried in the model specification stage, and often accomplished by the ad hoc method of adding increasingly unrealistic constraints to the model. One of the few programming model texts that even mentions model calibration is Hazell & Norton, who talk more of calibration tests than methods. This course will explicitly address these different stages of optimization model building and differs from the usual treatment by having a strong emphasis on model calibration.

Further Reading

- Judd.K.L. "Numerical Methods in Economics" MIT Press, Cambridge Mass.1999
Hazell P.B.R. & R.D Norton "Mathematical Programming for Economic Analysis in Agriculture" Macmillan Co New York, 1986
Howitt R.E. "Positive Mathematical Programming" American Journal of Agricultural Economics 77. (May 1995): 329 -342.
Paris Q. "An Economic Interpretation of Linear Programming" Iowa State University Press, Ames Iowa, 1991.
Paris, Q & R. E. Howitt. "An Analysis of Ill-Posed Production problems using Maximum Entropy" American Journal of Agricultural Economics. 80 (February 1998): pp 124-138.

II SPECIFYING LINEAR MODELS

Readings: Williams "Model Building . . ." Ch 3, pp. 20-47, Ch. 5, pp.63-82.
Hazell & Norton, Ch 2, pp. 9-31, Ch 3, pp. 32-53.

Constrained versus Unconstrained Models

Simple graphical models and nonlinear models in micro-economic texts are represented as unconstrained demand and supply functions and are optimized using calculus. A simple profit maximizing output is calculated given the following specification.

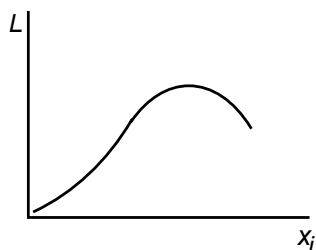
Given the general nonlinear production function $q = f(x_1, x_2)$, the price of the output q is p per unit output, and the cost per unit of input x_i is w_i .

If the objective is to maximize the profit Π subject to the production function $f(\cdot)$, the model is specified as:

$$\begin{aligned} \text{Max } \Pi &= pq - x_1 w_1 - x_2 w_2 \\ \text{subject to } q &= f(x_1, x_2) \end{aligned}$$

This equality constrained and differentiable problem can be expressed by the familiar Lagrangian function formulation, which by introducing the multiplier λ enables the equality constraint to be incorporated with the objective function. The resulting Lagrangian function can be optimized like an unconstrained function.

$$(A) \quad L = pq - \sum_i w_i x_i - \lambda (q - f(x_1, x_2))$$



The figure above represents the Lagrangian function, which can be maximized by the usual unconstrained approach of taking the partial derivatives $\frac{\partial L}{\partial x_i}$ and setting them equal to zero

(B) If the production function is defined as a linear relationship, defining a Leontief technology with fixed proportions of inputs per unit output, the problem becomes:

$$\begin{aligned} \text{Max } \Pi &= pq - w_1 x_1 - w_2 x_2 \\ \text{subject to } q - a_1 x_1 - a_2 x_2 &= 0 \end{aligned}$$

This linear profit maximizing production problem can be solved as a Lagrangian, and it can also be rewritten using linear algebra. Note that there is only one constraint in this example.

$$\begin{array}{ll} \text{Max} & c'x \\ \text{subject to} & a'x = 0 \end{array}$$

where $c' = [p, -w_1, -w_2]$
 $a' = [1, -a_1, -a_2]$

and the vector of input and output activities is: $x' = [q, x_1, x_2]$

Try multiplying this out to check that it is the same problem as above

Linear Programming

Linear Program: The equality constraint on the production function above is restrictive in that it implies that all the resources are exactly used up in the production processes. Given the nature of farm inputs such as land, labor, tractor time etc, inputs are available in certain quantities, but often they are not fully used up by the optimal production set. The relationship between the output levels q and the input levels x should be specified as inequality constraints for a more realistic and general specification. This inequality specification results in the Linear Program specification.

Given a set of m inequality constraints in n variables (x), we want to find the non-negative values of a vector x which satisfies the constraints and maximizes some objective function.

EXAMPLE

Yolo County Farm Model

In many places in this course we will use the following simple farm problem, based loosely on our local agriculture in Yolo County as a template to learn how linear programs are specified, solved and interpreted. We will use it for both analytical and empirical programming exercises.

The farm has the possibility of producing four different crops, alfalfa, wheat, corn and tomatoes. Yields are fixed so we can measure output by the number of acres of land allocated to each crop. The objective function is measured directly in net returns to the allocatable inputs. That is, the variable costs have been subtracted for simplicity. Constraints on production are all inequalities and represent the maximum amounts of land, water, and labor available, and a contract marketing constraint on the maximum quantity of tomatoes that the farmer can sell in any year.

The resulting linear program can be written as follows:

Maximize the scalar product of net returns $c'x$
 (Remember that c is $(n \times 1)$ and x is also $(n \times 1)$

Subject to the matrix of technical coefficients (A)
 and the vector of input resources available (b) $Ax \leq b$
 where (x) is the vector of production or activity levels.

The Yolo linear program is written as follows where the choice variables, measured in acres of land, are:

$$x_1 = \text{Alfalfa}, x_2 = \text{Wheat}, x_3 = \text{Corn}, x_4 = \text{Tomato}$$

The Objective Function is measured in dollars or other money units and maximizes

$$[121 x_1 + 160x_2 + 135x_3 + 825x_4]$$

Constraints

$$\begin{array}{l} \text{Land (acre)} \\ \text{Water (ac-ft)} \\ \text{Labor (hours)} \\ \text{Contract (tons)} \end{array} \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \\ 4.5 & 2.5 & 3.5 & 3.25 \\ 6.0 & 4.2 & 5.6 & 14.0 \\ 0.0 & 0.0 & 0.0 & 33.25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} 600.0 \\ 1800.0 \\ 5000.0 \\ 6000.0 \end{bmatrix}$$

The optimal solution is : $x^* = \begin{bmatrix} 0.0 \\ 419.549 \\ 0.0 \\ 180.451 \end{bmatrix}$

We expect tomatoes to come into the profit maximizing solution since they have a high profit margin per unit of land.

On one acre, we can grow 33.25 tons, but can only sell 6000 tons to the processor.

$\frac{6000.0 \text{ tons}}{33.25 \text{ tons/acre}} =$ a maximum acreage of 180.451 tomatoes. The rest of the land is used for the next most profitable crop, wheat. This optimal solution is not short of either water or labor.

The General Formulation of a Linear Program

		<u>Columns</u>					
<u>Row Name</u>		x_1	x_2	.	.	.	x_n RHS
Objective function		c_1	c_2	.	.	.	c_n
Resource constraints:							
1		a_{11}	a_{12}	.	.	.	$a_{1n} \leq b_1$
2		a_{21}	a_{22}	.	.	.	$a_{2n} \leq b_2$
		.	.				.
		.	.				.
		.	.				.
m		a_{m1}	a_{m2}	.	.	.	$a_{mn} \leq b_m$

NOTE: The previous problem can be written more compactly as:

$$\begin{array}{ll} \text{Max} & c'x \\ \text{Subject to} & Ax \leq b, \quad x \geq 0 \end{array}$$

Transformations

Mathematical precision is essential when formulating economic models for computers. We therefore need to think very precisely about the economic actions that we are trying to model. The most commonly modeled action are transformations. We will start with linear transformations since they are easier, but most micro theory is based on nonlinear transformations such as the decreasing utility that occurs when you eat too many donuts. All economic activities involve a transformation from input to output space or from product to utility space, such as eating donuts.

In our initial case of production, the economic transformation goes from goes from an “m dimensional” space of inputs, (b), to “n dimensional” space of outputs (x) and then to the scalar space of profit. In other words the production process being modeled takes a set of m inputs, say land, labor, and capital, and transforms them into n outputs, say corn, potatoes and milk, which are all sold for a common commodity, money. There are two transformations in this model of production. From inputs to outputs and from outputs to farm return. In addition we assume that the farmer is trying to maximize his return and will be constrained by some inputs. These simple transformations characterize the way in which most of the world’s population get their living. It is very important to be able to visualize the economic processes that underlie the linear algebra definitions, and be able to go back and forth between the algebraic definitions and the economic interpretations.

In the Yolo problem, the production transformation (mapping) is from land, water, labor, and contract constraints (m= 4) to alfalfa, wheat, corn, and tomatoes (n = 4) and the objective function transformation (mapping) is from 4-space to the 1-space of a single total farm return.

The mapping in Classroom space is:

$[\text{height, width, length}] \begin{bmatrix} 0 \\ 6 \\ 12 \end{bmatrix} = \text{scalar}$. The coordinates in 3 space locate a particular point on the floor that is 6 ft from one wall and 12 ft from another.

Other Examples of Production Transformations

$$\begin{array}{ccc} \text{Ford Cars} & [\text{Capital, Labor, Steel, Energy}] & \begin{bmatrix} a_{CT} & a_{CE} \\ a_{LT} & a_{LE} \\ a_{ST} & a_{SE} \\ a_{ET} & a_{EE} \end{bmatrix} = [\text{Expedition, Eclipse}] \\ & 1 \times 4 & 4 \times 2 \quad 1 \times 2 \end{array}$$

$$\begin{array}{ccc} \text{Candy} & [\text{Sugar, Chocolate, Gum, Corn Syrup}] & \begin{bmatrix} \text{milky} \\ \text{way} \\ \text{recipe} \end{bmatrix} = [\text{Milky Way Bar}] \\ & 1 \times 4 & 4 \times 1 \quad 1 \times 1 \end{array}$$

Linear Algebra Definitions Used in Linear Programming

Definition: Linear transformation: A linear transformation T from n space to m space is a correspondence on the space E^n which maps each vector x in E^n into a vector $T(x)$ in m -space. Transformations are performed by matrices or vectors as in the previous car or candy examples.

Note. Scalar multiplications can be carried through the **linear transformation**. That is, scalar multiplications can be factored out of the transformation as follows.

Example. Given the vectors x_1, x_2 , in, E_n , scalars λ_1, λ_2 , and the transformation $T(\cdot)$
 $T(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 T(x_1) + \lambda_2 T(x_2)$.

Definition: Linear dependence: If a vector a_i in a matrix A ($m \times n$) can be expressed as a linear combination of the other vectors then it is linearly dependent. (informal, intuitive definition)

Given the vectors $a_1, \dots, a_m$ from the space E^n , where $m < n$;

The vectors a_i are linearly dependent if there exist $\lambda_i, i = 1, \dots, m$

s.t. $\lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_m a_m = 0$ where not all the $\lambda_i = 0$. (math definition).

It is sometimes easier to see the opposite case of linear independence. In this case, if the vectors are linearly independent the only set of values for λ_i for which the linear transformation can be made to equal zero is when all the $\lambda_i = 0$.

An Example of Linear Dependence

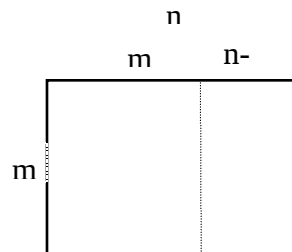
Set $l_1 = 1$.

From the definition of linear dependence, $l_1 a_1 + l_2 a_2 + \dots + l_m a_m = 0$,

Therefore setting $l_1 = 1$. $l_2 a_2 + \dots + l_m a_m = -a_1$

since a_1 is not $= 0$, then some $l_i, i = 2 \dots m$, must also be non zero.

Definition: The rank of a matrix is equal to the number of linearly independent vectors in the matrix.



Note 1. The number of linearly independent vectors cannot exceed m if $m < n$.

2. The number of linearly independent vectors cannot exceed smaller of the two dimensions, because rank is equal to the dimension of the largest invertible matrix. Remembering that matrices are only invertible if they are of equal dimensions (square)

The rank of A is denoted, $r(A)$

Existence of Solutions

Definition: Given a system of constraints $Ax = b$, a solution, the vector $\tilde{x}$, is a solution to this system if $\tilde{x}$ satisfies the constraints. We want to find the unique set of $\tilde{x}$ that optimizes the objective function value

Note: First: That the values of a solution vector x are the “weights” in a linear transformation.

Second: That since any set of values for x that satisfy the constraints is a solution, there is often a large number of potential feasible solutions, and the problem is to find the best one.

A Homogenous System

A system is defined as Homogeneous if all the values for the right hand side are zero ($Ax = 0$). In other words, b is defined as zero. Homogenous systems are often used as they are simpler to represent and manipulate. The trivial solution of $x = 0$ always exists. Non-homogenous systems can always be converted to homogenous systems by matrix augmentation.

Note We can convert $Ax = b$ to $\hat{A} \hat{x} = 0$, where $\hat{A} = [A \ b]$ the "augmented matrix" and $\hat{x} = \begin{bmatrix} x \\ -1 \end{bmatrix}$ the augmented vector

EXAMPLE

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

In the example above $Ax = b$, alternatively we can write the same equations in the form of $Ax - b = 0$.

$$\begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or redefining the matrices and vectors it equals $\hat{A} \hat{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Case 1. No Solution to the System Exists

No solutions exist to $Ax = b$ if the rank of A ($r(A)$) is less than the rank of the augmented matrix $\hat{A}$ where $\hat{A}$ is defined as the augmented matrix $[A \ b]$. Here we compare the rank for the augmented and unaugmented matrices.

The rank condition where $r(\hat{A}) > r(A)$ means that there are no solutions (except the trivial solution) since b is linearly independent of A .

Note that b must be linearly independent of all the vectors in A if augmenting A by b increases the rank of $\hat{A}$ over the rank of A .

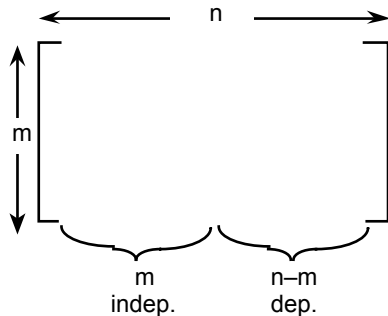
From the definition of a solution as a linear transformation, the linear independence of B from every vector in A means that no solutions can exist. In other words, the only set of weights (allocations) $\hat{x}$ that can make $\hat{A} \hat{x} = 0$ is the trivial solution when every value in $\hat{x} = 0$.

Case 2. The System has an Infinite Number of Solutions

For the system $Ax = b$, where A is an $m \times n$ matrix, and x is an $n \times 1$ vector.

If $r(A) < n$ then there exist an infinite number of non trivial solutions.

For simplicity set the rank of A to be $k = m$, ($n > m$) and arrange the linearly independent columns first, noting that the rank of $(A) = k$. The partitioned matrix dimensions are



shown:

Starting with the system $Ax = b$, where x is $n \times 1$, b is $m \times 1$, and A is $m \times n$, A and x can be partitioned as follows:

$$\text{Partition } A: \text{ into } [A_1 : A_2] \quad \text{and } x \text{ into } \begin{bmatrix} x_1 \\ \vdots \\ x_2 \end{bmatrix}$$

We can express the system $Ax = b$ as:

$$[A_1 \dots A_2] \begin{bmatrix} x_1 \\ \vdots \\ x_2 \end{bmatrix} = b \quad \text{or alternatively} \quad A_1 x_1 + A_2 x_2 = b$$

But by definition if $r(A) = m$, and A_1 has m linearly independent vectors then A_1 is "nonsingular" and A_1^{-1} exists. Rearranging and using the inverse yields

$$\begin{array}{lll} x_1 & = & A_1^{-1}(b - A_2 x_2) \quad \text{or multiplying it out yields:} \\ & = & A_1^{-1}b - A_1^{-1}A_2 x_2 \\ \text{solution} & & \text{known} \quad \text{known or zero} \\ \text{(unknowns)} & & \text{(chosen)} \end{array}$$

Note that the value of x_1 depends on x_2 , which can have an infinite number of values. This common situation leaves us with an infinite number of feasible solutions to search over for the optimal feasible solution. We make this intractable problem tractable by restricting our search to the finite number of feasible solutions that make up the basic solutions defined over the page.

Case 3. A Unique Solution Exists to the System

The system $Ax = b$ has a unique solution if:

(1) $r(A) = r(Ab)$

That is, the RHS augmented matrix has the same rank as A

(2) The matrix A is square and of full rank. That is, $A = m \times m$ and $r(A) = m$

SUMMARY

Solutions to the set of equations $Ax = b$ have the following conditions:

(a) $r(A) < r(Ab)$ No Solution

(b) $r(A) < \dim(x)$ Infinite Number of Solutions

**(c) $r(A) = \dim(x)$ Unique Solution
and A is full rank**

Basic Solutions

Definition: Given $Ax = b$, A is $m \times n$, $r(A) = m$, A basic solution to the system is when $(n - m)$ predetermined values of x (x_2) are set = 0.

Using the previous example, set $x_2 = 0$. The basic solution is $x_1 = A_1^{-1}b$

For convenience define A_1 as B, and A_2 as D. We can now write the system as:

$$x_B = B^{-1}b$$

where x_1 is denoted x_B , and x_2 (set = 0) is denoted x_D

Note 1. That since B is $m \times m$ and non-singular, the inverse B^{-1} exists

Note 2. Given that there are $(n-m)$ non basis vectors, there are many, but a finite number of alternative basic solutions.

Definition: Basic Feasible Solution (B.F.S.)

A basic feasible solution x is defined as:

Basic Solutions x such that $x = B^{-1}b$,

and Feasible

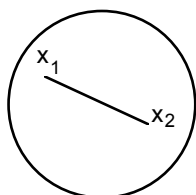
$$x \geq 0.$$

That is: a basic feasible solution that has non-negative values for the basic solution vectors.
A basic feasible solution where the x values in the basis are defined as all positive.

Definition: Convex Sets: A set $\{X\}$ is convex if for any points x_1 and $x_2 \subseteq \{X\}$, the line segment x_1, x_2 is also $\subseteq \{X\}$, i.e., "The set $\{X\}$ is convex if: There exist $x_1, x_2 \subseteq \{X\}$, such that the linear combination $\lambda x_1 + (1 - \lambda)x_2$ is also contained in $\{X\}$, for $0 \leq \lambda \leq 1$ "

This last line is interpreted as "anywhere on a line between x_1 and x_2 "

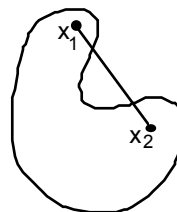
Note that $\lambda x_1 + (1 - \lambda)x_2$ is by definition a convex linear combination



$\lambda = 0$ implies that we are at point x_2

$\lambda = 1/2$ implies a point half way between x_1 and x_2

$\lambda = 1$ implies point x_1



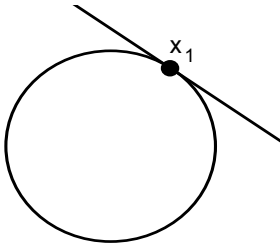
Intuition: Sets with holes or dents are not convex.

Definition: The Extreme Point of a convex set is x if $x \subseteq \{X\}$, but there do not exist any other $\bar{x}_1, \bar{x}_2 \subseteq \{X\}$ such that $x = \lambda \bar{x}_1 + (1 - \lambda)\bar{x}_2$ for $0 < \lambda < 1$.

What this says that an extreme point is a member of the set, but it cannot be expressed as a linear combination of any other points in the set. For an enjoyable empirical example, visit the lighthouse on the furthest west point of Point Reyes State park.

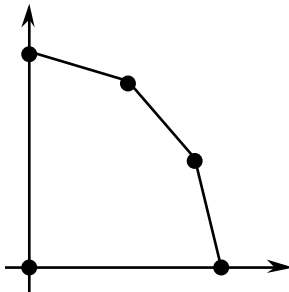
NOTE

- (1) In this definition λ is strictly $<$, not $\leq$
- (2) Any point that satisfies the definition above is a single point, because it is in the set, but only at an extreme point.
- (3) Intuitively, an extreme point of a convex set is part of the convex set, but cannot be expressed as a linear combination of any other two points in the convex set.



In a linear system an infinite number of solutions often exist, the objective function is used to select the maximum or minimum value. However to reduce the number of values to search for optimal value we use the properties of the basic feasible solution to reduce the search problem to one over a finite set of possible optimal values.

Basic feasible solutions are the non-negative extreme points.



The number of extreme points of a linear constraint set is finite. Accordingly, if we search the set of basic feasible solutions for the optimal value of the objective function, we will have found the optimal for the whole set.

Note that if the constraints are nonlinear, the resulting convex set now has an infinite number of solutions again. To solve for the optimum in this case we have to use a different approach that is addressed later in the course.

Slack or Surplus Variables

Slack or surplus variables in a linear program are used to convert the inequality constraints into equality constraints, thus making the problem easier to write mathematically and helping the interpretation of the model. They are sometimes called "artificial variables". While you have to understand the interpretation of these variables, in actual LP models they are usually put in automatically by the computer algorithm.

The two types of artificial variable correspond to the two types of inequality constraint. "Less than" ($\leq$) constraints are converted to equality constraints by **Slack** variables, while "Greater than" ($\geq$) constraints require **Surplus** variables.

Intuitively, if an inequality constraint may or may not be binding, but you want to always express it mathematically as a binding constraint there will be slack input to dispose of if the constraint is a "less than" ($\leq$) or a surplus to use up if the constraint is a "greater than" ($\geq$).

Example of a Slack variable in the Yolo model land constraint.

	Alfalfa	Wheat	Corn	Tomato	Slack Land	
Land	1	1	1	1	1	= 600

- Note:
1. Objective function values of slack/surplus variables are zero.
 2. Computer puts them in automatically (in GAMS).
 3. Gives us an initial basic feasible solution for the simplex method to use.
 4. The initial solution is:
 - (a) Guaranteed basic due to the diagonal constraint matrix
 - (b) Guaranteed feasible by the non-negativity condition on slack/surplus variables
 - (c) Won't add to the objective function value as they have zero objective function values

Surplus variables (for $\geq$ constraints) have a negative signed coefficient in the constraint because we want to reduce the surplus above the right hand side value, and thus reformulate the constraint as an equality.

Linear Program Objective Function Specification- Traditional Normative Approach

Economic Properties of Linear Program Objective Functions

- Linearity of the objective function in the parameters
(Max $c'x$ where c = the parameters)
- Constant Returns to Scale, i.e., cost/unit production is constant.
- Constant output prices (price-taker) (no regions, nations or large firms).

Some common examples of objectives are:

- Maximization of profit (Neoclassical Firm objective)
- Minimizing deviations from central planning targets
- Minimizing costs in a planned economy
- Minimizing risk of starving next season

Note: The units in the objective function are usually defined by the price units, for example \$ per ton. In the constraint matrix it is essential that there is consistency between the constraint units and objective function units.

EXAMPLE

Yolo Farm problem.

No costs are specified in the Yolo problem.

The objective function parameters are "Gross margins / acre." which are based on primary data and are equal to total revenue / acre – variable costs / acre.

This simple objective function specification works well until we need to consider capital investments, or changes in production technology as part of the problem to be solved. If changes in the amount of input used is part of the problem to be optimized then the net return per acre will change which requires a more complicated specification.

Specifying Linear Constraints

Types of Constraints

Linear constraints can be classified into three broad classes.

(a) Resource Constraints

In most linear models of production or distribution, resource constraints are intuitive. They usually take the form of a set of “m” summation constraints over the “n” activities. This form assumes that there are n possible production activities and m possible fixed resources used by the activities. The fixed resources are available in quantities $b_1 \dots b_m$.

The standard specification of this form is

$$\begin{array}{ll} \text{Max} & c'x \\ \text{subject to} & Ax \leq b \end{array}$$

The matrix A has individual elements, which are the input requirement coefficients for the production activities.

(b) Bounds

Where there are institutional limits on the activity levels, or because of bounds on the range of the linearity assumption, we may wish to bound the individual activity values. Bounds can be specified using a single row for each bounded activity. The general "less than or equal to" form of constraint can be used in the following way:

$$Ix \leq b$$

where the values for the b vector components b_i are the levels of the upper or lower bounds. Upper bounds have positive values for b_i . Lower bounds have negative values on both b_i and the corresponding 1 value in the identity matrix.

(c) **Linkage Constraints**

This type of constraint links two or more activities in a prespecified manner. The most common use of linkage constraints is to sum up total output or input use by activities. This operation is often needed where the total input use must be held in storage or purchased from another economic unit. Another common specification is an inventory equation that keeps track of commodity levels used, produced and on hand in the model. The units in the linkage constraint row usually determine the coefficients corresponding to each activity.

Linkage constraints are best approached systematically

- (1) Decide on the best units for the constraint row.
- (2) Write out the logic of the constraint in words. For example, a hay inventory row should be specified in tons. If the activities influencing the row are.
 - (i) Hay grown (acres)
 - (ii) Cattle to be fed (head)
 - (iii) Hay bought (tons)
 - (iv) Hay on hand in storage (tons)

Constraint logic – "Hay consumed by cows plus the change in storage equals hay grown plus hay purchased." The easiest way to think of linkage constraints is to define the "Flows in" and "Flows out" of the commodity that the constraint defines. Then decide if the problem requires that the flows in are greater than, less than, or equal to the flows out.

- (3) Although linkage constraints are usually equalities, LP problems solve more easily if the constraints are specified as inequalities. The trick is to specify the signs of the coefficients so that the constraint is driven to hold as an equality by the objective function. For example if a constraint is set as a "greater than" inequality that requires a basic ration for animal production but allows a greater amount of food to be fed. An optimizing model will always constrain the ration to the basic level if the food input is costly or constrained.

Example 1

Maximizing net revenue. Three output activities x_1, x_2, x_3 .

All activities require different quantities of a variable

input x_4 , with a_{ij} input requirement coefficients of : a_{41}, a_{42}, a_{43} .

Objective Row	c_1	c_2	c_3	$-c_4$
Linkage Row	a_{41}	a_{42}	a_{43}	$-1 \leq 0$.

The row is measured in units of the variable input x_4 . The cost per unit is c_4 . The linkage row above allows the possibility that more x_4 can be purchased than is needed by production

activities x_1 , x_2 and x_3 . However, since the extra units have a cost associated with them the objective function will be reduced by the “slack” activities and hence an optimizing model will not purchase more input than is needed. The linkage row does require that the units of x_4 required by the production of x_1 , x_2 and x_3 are summed and that this sum is less than or equal to x_4 . In short, you can buy too much x_4 , but you cannot buy too little.

(4) If the problem solution is unbounded, check the signs in the constraint.

(5) Check the operation of the linkage constraint by hand calculations on a representative constraint at the optimal solution values.

(6) Inventory stocks can be incorporated most simply by right hand side values, (example 2). An alternative approach is to specify a separate column for the inventory stock activity, which may itself be constrained by upper or lower bounds (example 3). This approach is required if the stock is priced as a separate activity in the objective function.

Example 2

A stock of 50 units of x_4 is available at the start of the problem.

Objective Row	c_1	c_2	c_3	$-c_4$	
Linkage Row	a_{41}	a_{42}	a_{43}	-1	≤ 50 .

The resulting constraint in the optimal solution will have the form:

$$x_1 a_{41} + x_2 a_{42} + x_3 a_{43} - x_4 \leq 50.$$

The interpretation is that the model has to satisfy the requirements of activities $x_1 \dots x_3$ by using some of the unpriced 50 units on hand, or they can buy inputs for a price c_4 through activity x_4 .

Example 3

The problem starts with no stocks of x_4 , but is required to have 100 units in stock at the optimal solution. We now specify activity x_5 as the stock of input x_4 , for instance hay. Activity x_4 is defined as purchasing or growing hay.

Objective Row	c_1	c_2	c_3	$-c_4$	c_5	
Linkage Row	a_{41}	a_{42}	a_{43}	-1	1	≤ 0
Minimum Stock Row					-1	≤ -100 .

Models of Transportation Problems

Linear optimization is particularly good at solving problems that minimize the cost of transporting a commodity from defined sources to destinations.

If there are:

n sources defined by $i = 1 \dots n$

m destinations defined by $j = 1 \dots m$,

Note- there are n times m possible ways to transport the commodity.

The activity (the amount shipped) is therefore defined as x_{ij} and has an associated cost of transport of c_{ij} .

The objective function is therefore to Minimize $z = \sum_i \sum_j c_{ij} x_{ij}$

Demand at destinations.

The transport problem is constrained by a set of minimum demand quantities at each destination.

If the quantity demanded at destination j is defined as b_j the demand constraint is :

$$\sum_i x_{ij} \geq b_j$$

It says

“ the sum of the amounts that arrive from all sources must be greater or equal to ($\geq$) the amount demanded at destination j ”

Source “Supply” Constraints

The total amount shipped out of any supply source cannot exceed its capacity. Given a maximum capacity of a_i at source i , the supply constraint is :

$$\sum_j x_{ij} \leq a_i$$

It says

“The amount shipped from source i to all destinations must be less than or equal to ($\leq$) the amount available at source i ”.

The complete transportation problem for n sources ($i = 1 \dots n$) and m destinations ($j = 1 \dots m$) is:

$$\text{Minimize Total Cost } z = \sum_i \sum_j c_{ij} x_{ij}$$

$$\text{subject to } \sum_j x_{ij} \leq a_i$$

$$\sum_i x_{ij} \geq b_j$$

Demand Shortage Costs and Transportation Cost

Often problems can be written more realistically by redefining the demand constraints as not having to hold exactly, but to incur “Shortage Costs” if they are not met. To influence the optimal solution, the outage costs of not having enough product must exceed the transportation and supply costs.

Outage activities are included in the left-hand side of the demand constraints.

$$\sum_i x_{ij} + \text{out}_j \geq b_j$$

Given an shortage cost of "cout_j", the transportation model objective function is now

$$\sum_i \sum_j c_{ij} x_{ij} + \sum_j \text{cout}_j \text{out}_j$$

The model now finds the optimum pattern of transportation for the supplies on hand, and calculates the cost minimizing way to spread the shortage among destinations.

III SOLVING LINEAR MODELS

Solution Sets

A linear equality constraint defines a line in two space, a plane if it is in three space, and a hyperplane if the constraint is in n dimensions. It follows that a linear constraint (inequality) in “n space” divides the space into two half spaces. Therefore the set of values that satisfy several linear constraints must be common to (or contained in) the intersection of several half-spaces. Fortunately it turns out that the intersection of linear half-spaces is a convex set. Therefore the set of possible solutions which can satisfy several linear constraints at the same time is a convex set. This convex set is known as the feasible solution set for the linear inequality constraints, since any point in this set can satisfy all the constraints. We use the properties of convex sets to search over the large set of possible solutions in an efficient way for the optimal solution that maximizes some particular objective.

The Fundamental Theorem of Linear Programming

Given a linear programming problem in the usual matrix algebra form:

$$\begin{array}{ll} \max & c'x \\ \text{subject to.} & Ax \leq b \end{array}$$

where A is m x n with rank (A) = m:

The Theorem can be summarized by stating that:

1. If a feasible solution exists to the problem, a basic feasible solution to the problem also exists.
2. If an optimal feasible solution exists, then an optimal basic feasible solution to the problem also exists.

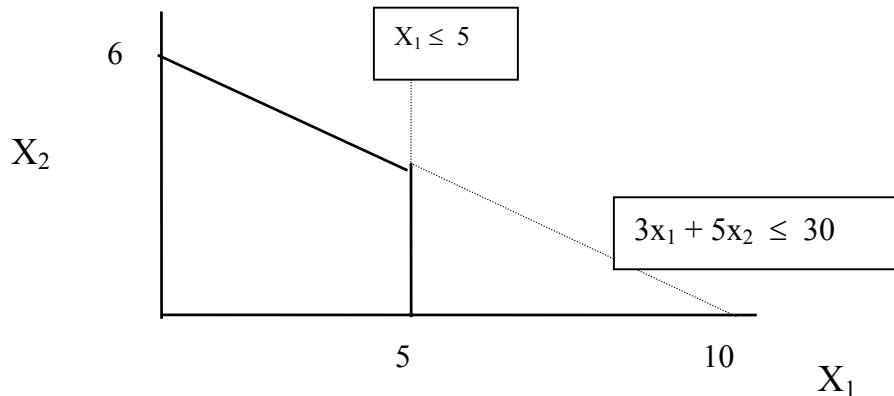
Therefore all we have to do is check which of the basic feasible solutions maximizes the objective function and we know that we have checked all the possible candidates for the optimal basic feasible solution.

EXAMPLE

Given the following set of inequality constraints:

$$\begin{array}{rclcl} x_1 & & & \leq & 5 \\ 3x_1 & + & 5x_2 & \leq & 30 \\ x_1, x_2 & \geq & 0 & & \end{array}$$

We can show these inequality constraints graphically in x_1, x_2 space.



The intersection of the two half-spaces is the feasible solution space. Note that the feasible solution set is a convex set with four extreme points at each corner.

Extreme Points and Basic Solutions

Given system $Ax = b$, $x \geq 0$

where A is $m \times n$, and the rank of A , $r(A) = m$:

If K is the convex set of all solutions to the system (i.e., the set of possible $(n \times 1)$ vectors x that satisfy the system).

Then a vector x is an extreme point of K if x is a basic solution to the system.

Definition: A Basic solution: is defined as a solution where all the basic variables are non-zero, all other non-basic variables are zero.

Corollary: If a feasible solution exists then there exist a finite number of potentially optimal solutions.

Given that we can be sure that our optimal solution is among the basic feasible solutions we can now concentrate the search for optimal solutions among the finite set of basic solutions. Note that for the feasible solution set, the number of extreme points equals the numbers of binding constraints which also equals the number of non-zero basis activities.

An Introduction to the Simplex Algorithm

The Simplex algorithm

An algorithm is a set of systematic instructions to the computer that enables us to program it to perform a given task.. The Simplex algorithm is one of the oldest but still the best algorithm for most problems of linear optimization. Its operation can be summarized as:

- 1 – Changing the basis of the problem, and hence the solution, by changing basis vectors
- 2 – Using the objective function value for a systematic choice of basis vectors that always improve the objective function.

Since the algorithm is driven by the effect of a change of basis on the objective function, to understand its operation we need to analyze the algebra and economics of a change of basis for the following familiar LP problem.

$$\begin{aligned} \text{Max } P &= c'x \\ \text{s.t. } Ax &= b, \quad x \geq 0, \quad A \text{ is } m \times n, \quad \text{and rank } A = m. \end{aligned}$$

Partition A into basis and non-basis matrices denoted respectively B and D.

Defining the partition of $A \rightarrow [B \mid D]$

$$\begin{array}{cc} \neq & \neq \\ \text{basis} & \text{non-basis} \\ m \times m & m \times (n-m) \end{array}$$

when partitioned, x and c' become: $x = \begin{pmatrix} x_B \\ x_D \end{pmatrix}$ $c' = \begin{pmatrix} c'_B & c'_D \end{pmatrix}$

The L.P. problem becomes: $\text{max } \begin{pmatrix} c'_B & c'_D \end{pmatrix} \begin{pmatrix} x_B \\ x_D \end{pmatrix}$ s.t. $\begin{pmatrix} B & D \end{pmatrix} \begin{pmatrix} x_B \\ x_D \end{pmatrix} \leq b$

or multiplying out the partitioned matrices results in:

$$\begin{aligned} \text{Max } P &= c'_B x_B + c'_D x_D \\ \text{s.t. } Bx_B + Dx_D &\leq b \\ x_B &\geq 0, \quad x_D \geq 0 \end{aligned}$$

An optimal basic feasible solution (BFS) has the elements of x_B are all non-zero (assuming non-degenerate solutions, dealt with later), and all the x_D elements are zero. The constraints for a basic feasible solution become:

$$Bx_B + D [0] = b \quad \text{or more concisely,} \quad Bx_B = b$$

Since B^{-1} exists by definition of a basis ($m \times m$, m linearly independent rows)
A **basic feasible solution** to the system is:

$$\underline{x_B = B^{-1} b}$$

However the point of this analysis is to find the effect of a change of the basic solution on the value of the objective function. Accordingly, the partitioned x vector is used to write the objective function in terms of basis and non basis variables.

The partitioned objective function is $\Pi = c'_B x_B + c'_D x_D$ or substituting in the solution above:

$$\Pi = c'_B B^{-1}b + c'_D x_D \text{ since } x_D = 0$$

For basic solutions the objective function is: $\Pi = c'_B B^{-1}b$

What happens to the objective function value (z) when we consider introducing one of the x_D (non basis vectors), say " x_j " into the basis? The analogy is the "In group" and "Out group" in US High schools where teenage popularity is very important. One way in which one can judge and be judged as to what group you are in is with whom you eat lunch. Assume for mathematical reasons that the lunch bench has a finite dimension (the rank of the basis matrix) and the number of people who can sit on the bench is limited. Since teenage popularity is fickle, it is quite likely that individuals (vectors) move in and out of the "In group" over time. The point is that if a new individual is popular enough to be admitted to the "In group", someone will have to leave the bench (the basis) and the other members of the group will have to rearrange their seating on the bench with the arrival of the new entrant. Mathematically this process can be represented as:

Problem: Those who stay have to move over, and someone (a vector) will have to leave the basis. Writing out a solution to the partitioned constraints yields:

$$Bx_B + Dx_D = b \quad \text{subtracting the non basis values}$$

$$Bx_B = b - Dx_D \quad \text{premultiplying by } B^{-1}$$

$$x_B = B^{-1}b - B^{-1}Dx_D$$

Now we substitute this result into the objective function to see the effect of the introduction of a non-basis activity into the basis. This changes the objective function value.

Since $\Pi = c'_B x_B + c'_D x_D$ substituting in for x_B from above yields:

$= c'_B(B^{-1}b - B^{-1}Dx_D) + c'_Dx_D$ collecting the terms multiplied by x_D
we can now factor out x_D to get:

$$\Pi = \underbrace{c'_B B^{-1}b}_{\text{original z value}} + \underbrace{(c'_D - c'_B B^{-1}D)}_{\substack{\text{vector of} \\ \text{revenues} \\ \text{from the} \\ \text{new activity} \\ \text{c}_j \text{ value}}} \underbrace{x_D}_{\substack{\text{vector of the cost of} \\ \text{moving the } x_B \text{ basis to} \\ \text{accommodate the new} \\ x_j \text{ vector} \\ z_j \text{ value}}}$$

Note.

$(c'_D - c'_B B^{-1}D)$ is a vector representing net values of x_D 's (non basis vectors). In the basic equation for Π they're multiplied by vector x_D of zeros. However if one of the x_D values is set to non zero, that is, it is brought into the basis, the objective function will be changed by this amount.

The net change in objective function value from moving a new vector in is:

$$\underbrace{(c'_D - c'_B B^{-1}D)}_{\substack{\text{revenue contributed} \\ \text{by a new activity} \\ \text{from } x_D}} - \underbrace{c'_B B^{-1}D}_{\substack{\text{change in the } x_D \text{ values} \\ \text{forced to satisfy the} \\ \text{constraints defined as: } y_j \\ \text{times the unit revenue lost}}}$$

In other words

$$\text{Marginal new revenue} - \text{Marginal opportunity cost}$$

That is $(c'_B B^{-1}D)$ is the cost of moving the old basis values to fit an x_D vector. In other words, $(c'_D - c'_B B^{-1}D)$ is the (revenue – opportunity cost) of vector x_D , the incoming non basic vector.

If you want to maximize $c'x$, you select the change of basis that maximizes the difference between revenue and opportunity cost of the new vector. That is, for a maximization problem we select among those vectors that have:

$$\underline{c'_D} > \underline{c'_B B^{-1} D}$$

Using this criterion we only change the basis if we improve the objective function.

The **Dual price** or “**shadow value**” is defined as

$$\underline{\lambda' \equiv c'_B B^{-1}}$$

Note that λ is the “marginal” associated with the constraint rows in the Gams printout.

Substituting the expression for λ above into the equation defines the term z_j where:

$$z_j = [c'_B B^{-1} D]_j = \lambda' D$$

The Reduced cost is defined as:

$$r_j = c_j - z_j \equiv (c'_D \geq c'_B B^{-1} D)$$

For a **maximization** problem, the algorithm rule is:

If there are any r_j greater than zero ($r_j > 0$), add the vector with the highest r_j .

If all $r_j \leq 0$, you're at optimal solution.

In a **minimization** problem, the rule is to bring in the x vector with lowest r_j .

If r_j 's are all ≥ 0 , the minimization problem is at the optimum.

Intuition.

r_j is the net benefit of an activity entering the basis and is the "marginal" on the **variables** in the GAMS printouts. See the Rock Music example in chapter XI

c_j is the “benefit” of activity j (the incoming activity) to the objective function

z_j is the opportunity cost of moving the current basis values to accommodate incoming unit of x_j , and is equal to the lost revenues from current basis activities.

$$z_j = \begin{array}{ccc} c'_B B^{-1} D & = & \lambda' D \\ \uparrow \quad \uparrow & & \uparrow \\ \text{revenue from} & \text{input release} & \text{shadow values} \\ \text{x's in basis} & \text{quantities for} & \text{of resources b times} \\ & \text{x's} & \text{the input requirement} \end{array}$$

Alternatively the same z_j value can be expressed **in terms of the y_j vector ,where y_j is defined as $B^{-1}d_j$**

$$z_j = c'_B \underbrace{B^{-1}d_j}_{y_j}$$

$\uparrow$ value (revenue) of a unit of each of those x_i s $\uparrow$ y_j = The reduction in the output of the current basis activities x_i needed to release resources for 1 unit of the entering x_j

Π = value of objective function

$$\Pi = c'x = c'_B x_B + c'_D x_D = c'_B x_B \quad \text{therefore can also be written as:}$$

$$\Pi = c'_B \underbrace{B^{-1} b}_{x_B}$$

An Outline of the Simplex Method

The Simplex algorithm optimizes using four critical pieces of information.

(i) The **Reduced costs** of the non basis activities, defined in the vector r_j :

$$(1) \quad \begin{matrix} [c'_B B^{-1} D - c_D] = & z_j - c_j = r_j. \\ \uparrow \quad \quad \uparrow & \quad \quad \uparrow \quad \uparrow \\ \text{cost} \quad \text{revenue} & \quad \quad \text{cost} \quad \text{revenue} \end{matrix}$$

Note: The sign of r_j is reversed here - always check the LP package for the definition of r_j .

(ii) The value of the **current Objective function**

$$(2) \quad c_B B^{-1} b = c_B x_B = \Pi$$

Recall $\Pi = c x = c_B x_B + c_D x_D$, but $c_D x_D = 0$ for a basic solution, since x_D equals zero.

(iii) The **values of a new parameter** y_j that is calculated for all non basis vectors that may enter the basis at the next iteration.

$$(3) \quad B^{-1} D = [y_1, y_1 \dots y_{n-m}] \quad B^{-1} D \text{ is an } m \times (n-m) \text{ matrix.}$$

y_j is a $(m \times 1)$ column vector of resource requirements y_{ij} . y_{ij} = the amount of resource i used in the current basis, and needed by a unit of x_j (if it enters basis). That is, the amount x_i in the basis must "move over" to accommodate x_j entering basis.

(iv) The **value of the current basis variables** x_B is given as:

$$(4) \quad B^{-1} b = x_B$$

An overview of the simplex method builds on the four matrix expressions derived above.

Step I. First Iteration. Your GAMS/MINOS adds slack or surplus variables to convert all the inequality constraints to equalities. The slacks have two important characteristics

- (i) Zero values in the objective function coefficient row (c)
- (ii) A constraint matrix which is an identity matrix.

Thus, an initial basis of all slack and surplus variables is always

- (i) Of full rank
- (ii) Feasible
- (iii) Zero value objective function and therefore, zero opportunity costs z_j for alternative activities.

Step II. For each x_j not in the basis calculate the **reduced cost** $r_j = (c_j - z_j)$, which is the net benefit of x_j **entering the basis**. (For the first iteration, $r_j = c_j$). Select the x_j with maximum $r_j > 0$ to enter the basis. (For a maximization problem).

For a maximization problem - if any $r_j > 0$, the problem is not optimal, therefore, select the maximum $r_j > 0$ to enter the basis.

Step III. The y_j vector is used to **select the activity that leaves the basis**.

y_j is an $m \times 1$ vector of values that show the rate of substitution between the basis activities x_{B_i} and the incoming vector x_k chosen in step II

From equation (3) we see that $B^{-1}D$ yields $(n - m)$ " y_j " vectors, each with m elements. If we pick a non basis vector in D , say d_k , to enter the basis, the corresponding y_k vector will be:

$$(5) \quad y_k = B^{-1}d_k = B^{-1}a_k$$

Using (5) we can also write a_k - the input requirements for the incoming vector- as a function of y_k :

$$(6) \quad a_k = By_k$$

But since B is the basis submatrix of A ,

$$(7) \quad B = [a_1, a_1, \dots, a_m]$$

and thus from (7) the potential incoming vector a_k selected in step II can be expressed as a linear combination of the basis vectors. Substituting (7) back into (6) yields:

$$(8) \quad a_k = a_1 y_{1k} + a_2 y_{2k} + \dots + a_m y_{mk}$$

we can also multiply each element in (8) by an arbitrary nonzero scalar ε

$$(9) \quad \varepsilon a_k = a_1 \varepsilon y_{1k} + a_2 \varepsilon y_{2k} + \dots + a_m \varepsilon y_{mk}$$

Expanding the basic feasible solution $Bx_B = b$ using (9) for B results in.

$$(10) \quad a_1 x_1 + a_2 x_2 + \dots + a_m x_m = b$$

We now introduce $-\varepsilon a_k$ and εa_k into the feasible basis in equation (10).

Note (1) If $\varepsilon = 0$, no basis change occurs.

(2) The algebraic trick of adding and subtracting within an identity

$$(11) \quad a_1 x_1 + a_2 x_2 + \dots + a_m x_m - \varepsilon a_k + \varepsilon a_k = b$$

Multiplying (9) by -1 and substituting the right hand side of (9) into (11) for $-\varepsilon a_k$ yields

$$(12) \quad a_1 x_1 + a_2 x_2 + \dots + a_m x_m + \{-a_1 \varepsilon y_{1k} - a_2 \varepsilon y_{2k} \dots - a_m \varepsilon y_{mk}\} + \varepsilon a_k = b$$

factoring out the $a_1, \dots, a_m$ vectors gives

$$(13) \quad a_1(x_1 - \varepsilon y_{1k}) + a_2(x_2 - \varepsilon y_{2k}) + \dots + a_m(x_m - \varepsilon y_{mk}) + \varepsilon a_k = b.$$

Point. As we increase the value of the scalar ε , the influence of a_k increases, and $(x_i - \varepsilon y_{ik})$ will be driven to zero for some activity i. This activity will leave the basis.

Note

1. That by the definition of a basic solution, when an activity is driven to a zero value it leaves the basis.
2. An $m + 1$ dimensional object in m dimensional space is called a SIMPLEX, hence the name of this algorithm.
3. The first term to be driven to zero will have the smallest x_i / y_{ik} since $(x_i - \varepsilon y_{ik}) = 0$ when $\varepsilon = x_i / y_{ik}$.

From (13) we can see two outcomes of changes in the value of ε :

1. If $\varepsilon = 0$ we get the old basis in (7)
2. If ε is made large, the importance of the new vector a_k increases, but there is a danger that one of the new variable values $(x_i - \varepsilon y_{ik})$ will be driven to a negative value, which would produce an infeasible solution.

Question – How do we select the value of ε that will drive one basis activity level to zero, without driving any others to negative values which would make them infeasible ?

$$(14) \quad x_i - \varepsilon y_{ik} = 0 \Rightarrow \varepsilon = x_i / y_{ik}$$

Point: If we pick the exiting vector to be the first basis vector to have its coefficient driven to zero by the entry of a_k in the basis, we will have a new basis and ensure feasibility.

The criterion is therefore $\text{Min} \left\{ \frac{x_{Bi}}{y_{ij}} : y_{ij} > 0 \right\}$, NOTE x_{Bi} and y_{ij} are scalars

If no $y_{ij} > 0$, the problem is unbounded. Essentially it says that the rate of trade off between the inputs is negative.

Why: If resource requirement $y_{ij} < 0$ then adding x_j to basis will free up resources necessary for x_i (but may consume resources necessary for other activities). If $y_{ij} < 0$ for all i then adding x_j to basis frees up resources and doesn't consume any. So you do it forever.

Step IV. Proceed with these iterations (return to Step II) changing the basis each time until (for a maximizing problem) all $r_j \leq 0$. You now have the optimal solution. $X_B = B^{-1}b$ and objective function $\Pi = c_B X_B$

An Empirical Example of the Matrix Simplex Solution

As an illustration of matrix manipulations involved in solving for solutions of systems of linear equations, we develop the problem of the music (CD) production firm who can promote bands in four groups XA (Alternative Rock) , XC (Country) , XG (Grunge Rock) , XH (Hip Hop)

Assume that the firm has two stocks of input needed to make a CD a successful seller, namely promotion Airplay time (AP) and recording studio time (ST). Both these assets are fixed in their maximum availability, max AP = 620, max ST = 180 In addition, the music firm manager knows how much of each input is required for each type of band. The technology required for the music business can therefore be represented by the following set of equations. $Ax \leq b$.

$$\begin{array}{cccccc} XA & XC & XG & XH & & \\ 25 & 32 & 18 & 28 & \leq & 620 \\ 12 & 14 & 17 & 10 & \leq & 180 \end{array}$$

To convert the set of inequality constraints into a set of equality equations we add two more activities for the slacks on AP and ST, respectively S1 and S2.

$$\begin{array}{ccccccc} XA & XC & XG & XH & S1 & S2 & \\ 25 & 32 & 18 & 28 & 1 & & = 620 \\ 12 & 14 & 17 & 10 & & 1 & = 180 \end{array}$$

The manager also knows the gross margin for each type of CD. Under current market conditions they are:

$$\begin{array}{cccccc} CA & CC & CG & CH & S1 & S2 \\ 3.5 & 4.2 & 5.6 & 4.8 & 0 & 0 \end{array}$$

The simplex method starts the search for the optimal solution with basic solution that we know will always be feasible and able to be improved on. The initial basic solution is composed of the slack variables for the binding constraints. In this example the initial basic solution is composed of the vectors S1 and S2. Therefore the initial basis called B_1 is :

$$B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ with the inverse being the same matrix.}$$

$$\text{The basic solution } X_{B1} = B_1^{-1} b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 620 \\ 180 \end{bmatrix} = \begin{bmatrix} 620 \\ 180 \end{bmatrix}$$

Since the gross margin from slack resources is zero the objective function is also zero.

The value for the objective function is:

$$\Pi_1 = c_{B1}' X_{B1} = \begin{bmatrix} 0.0 & 0.0 \end{bmatrix} \begin{bmatrix} 620 \\ 180 \end{bmatrix} = 0.0$$

To select the next activity to come into the basis we have to calculate the vector of r_j ($c_j - z_j$) values for the four music activities that are currently in the non basic set X_{D1} .

Since the formula for the vector of z_j values is:

$z = c_{B1}' B_1^{-1} d$ and since c_{B1} is composed of zero values, the opportunity cost of using slack inputs to produce CDs is zero. Thus the value of the vector of r_j s is equal to c_{D1} .

$$r_{j1} = \begin{bmatrix} 3.5 \\ 4.2 \\ 5.6 \\ 4.8 \end{bmatrix}$$

Since we wish to increase the objective function as fast as possible we select the largest r_j value, which brings XG (Grunge Rock) into the basis. To calculate the level at which we can bring in the Grunge rock band and which slack activity leaves the basis, we now calculate the y_j vector for XG.

$$y_{XG} = B_1^{-1} d_{XG} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 18 \\ 17 \end{bmatrix} = \begin{bmatrix} 18 \\ 17 \end{bmatrix}$$

Applying the criteria for the exiting activity also sets the level at which the new activity comes into the basis.

$$\text{Min} \left\{ \frac{X_{Bi}}{y_{iXG}} : y_{iXG} > 0 \right\} = \text{Min} \left\{ \frac{620}{18}, \frac{180}{17} \right\} = \text{Min} \left\{ 34.44, 10.58 \right\}$$

Accordingly the Grunge rock comes in at a level of 10.58 units which drives the slack on recording studio time to zero, that is out of the basis. The new basis B_2 is composed of S1 and XG activities and is:

$$B_2 = \begin{bmatrix} 1 & 18 \\ 0 & 17 \end{bmatrix} \text{ and } B_2^{-1} = \begin{bmatrix} 1 & -\frac{18}{17} \\ 0 & \frac{1}{17} \end{bmatrix}$$

The new solution for the basic activities is:

$$X_{B2} = B_2^{-1} b = \begin{bmatrix} 1 & -\frac{18}{17} \\ 0 & \frac{1}{17} \end{bmatrix} \begin{bmatrix} 620 \\ 180 \end{bmatrix} = \begin{bmatrix} 429.41 \\ 10.588 \end{bmatrix}$$

The new value for the objective function is:

$$\Pi_2 = c_{B_2}' X_2 = [0.0 \quad 5.6] \begin{bmatrix} 429.41 \\ 10.588 \end{bmatrix} = 59.29$$

This level of return is clearly better than the initial solution of not using the resources at all, but is it the best use that we can make of the limited studio resources ?

With the new basis there is a new set of opportunity costs for the resources.

Studio space is fully used on the Grunge band under the current allocation, but Airplay time still has a lot of slack.

The new set of r_j values are as follows:

$$\begin{aligned} r &= c_{D_2} - z_{D_2} = c_{D_2} - c_{B_2}' B_2^{-1} D_2 \\ &= c_{D_2}' - [0 \quad 5.6] \begin{bmatrix} 1 & -\frac{18}{17} \\ 0 & \frac{1}{17} \end{bmatrix} \begin{bmatrix} 253228 \\ 121410 \end{bmatrix} = [-0.45 \quad -0.41 \quad 1.51] \end{aligned}$$

Using the maximum r_j rule, the only music activity for which the marginal contribution to the objective function exceeds that of the Grunge band is XH (Hiphop). Therefore XH comes into the basis.

The new y_j values are:

$$y_{XH} = B_2^{-1} d_{XH} = \begin{bmatrix} 1 & -\frac{18}{17} \\ 0 & \frac{1}{17} \end{bmatrix} \begin{bmatrix} 28 \\ 10 \end{bmatrix} = \begin{bmatrix} 17.412 \\ 0.588 \end{bmatrix}$$

Applying the criteria for the exiting activity also sets the level at which the new activity comes into the basis.

$$\text{Min} \left\{ \frac{X_{B_i}}{y_{iXH}} : y_{iXH} > 0 \right\} = \text{Min} \left\{ \frac{429.41}{17.412}, \frac{10.58}{0.588} \right\} = \text{Min} \left\{ 24.66, 17.99 \right\}$$

This says that the Grunge band should give up their studio time to the Hiphop band, and airplay time will still be slack. The next (third) basis has S1 and XH and is:

$$B_3 = \begin{bmatrix} 1 & 28 \\ 0 & 10 \end{bmatrix} \text{ and } B_3^{-1} = \begin{bmatrix} 1 & -2.8 \\ 0 & 0.1 \end{bmatrix}$$

The new solution for the basic activities is:

$$X_{B_3} = B_3^{-1} b = \begin{bmatrix} 1 & -2.8 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 620 \\ 180 \end{bmatrix} = \begin{bmatrix} 116.0 \\ 18.0 \end{bmatrix}$$

The new value for the objective function is:

$$\Pi_3 = c_{B_3}' X_{B_3} = [0.0 \quad 4.8] \begin{bmatrix} 116.0 \\ 18.0 \end{bmatrix} = 86.4$$

Clearly Hiphop is an improvement over the Grunge bands.

The r_j values for the new basis B_3 are as follows:

$$r = c_{D3} - z_{D3} = c_{D3} - c_{B3}' B_3^{-1} D_3 = c_{D3} - [0 \quad 4.8] \begin{bmatrix} 1 & -2.8 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 25 & 32 & 18 \\ 12 & 14 & 17 \end{bmatrix} = [-2.26 \quad -2.52 \quad -2.56]$$

Since all the r_j values for the third basis are negative this tells us that we are at the optimum solution with the Hiphop band using all the Studio time and Airplay time is in surplus.

This problem is specified and solved in Gams over the page. Check the format of the Gams set up in matrix form. Note that every number on the Gams output has been, or can be, calculated in the matrix operations by hand. Remember that the “Duals” or “Marginals” on the resource constraints are calculated above in the r_j equation as

$$\lambda = c_{B3}' B_3^{-1} .$$

The Gams code to solve this problem and the resulting optimal solution for the problem can be downloaded from the Gams template part of the class webpage. Check that the matrix calculations agree relatively closely with the Gams printout.

IV The DUAL PROBLEM

The problem of minimizing the cost of inputs subject to constraints on a minimum output level is equivalent to the problem of maximizing profit subject to production technology and constraints on the total input available. Thus every optimization problem can be posed in its Primal or Dual form. For every Primal Problem there exists a Dual Problem which has the identical optimal solution. So far in this course we have only dealt with the primal form of the problem since its intuitive explanation is easier. The standard form of the twin Primal and Dual problems is written as follows:

Primal

$$\begin{aligned} \text{Max } c'x \\ \text{s.t. } Ax \leq b \\ x \geq 0 \end{aligned}$$

Dual

$$\begin{aligned} \text{min } \lambda'b \\ \text{s.t. } A'\lambda \geq c \\ \lambda \geq 0 \end{aligned}$$

where:

x is $(n \times 1)$ vector of primal variables and λ is $(m \times 1)$ vector of dual variables

The Objective functions ask the following questions:

Primal

What is the maximum value
of firm's output?

Dual

What is minimum acceptable price
that I can pay for the firm's assets ?

The dual specification of a problem is particularly useful:

1. When the Dual specification is simpler to solve than the Primal specification
2. When you know production costs but not production technology
3. When it's the dual values that interest you (often for Economists)

The Economic Meaning of the Dual

The Dual Variables λ_i are elements in the vector of imputed marginal values of the resources b_i

Equivalent intuitive interpretations are the opportunity costs of not having the last unit of resource, or equivalently, how much you'd pay for one more unit of the resource b_i

$$\lambda_i = \text{imputed value of } b_i = \frac{\partial(\text{obj})}{\partial b_i} \geq 0 \quad \text{and can be thought of as the marginal effect on the objective function of a small change in resource availability.}$$

Note: If the constraint isn't binding λ_i is always equal to zero, by the Kuhn Tucker complementarity slackness conditions addressed in detail later .

Dual Objective function

The dual objective function $\lambda' \mathbf{b}$ is equal to the sum of the imputed values of the total resource stock of the firm. It is the sum of money that you would have to offer a firm owner to make them consent to a buy-out.

Dual Constraints

The dual constraints $\mathbf{A}'\lambda \geq \mathbf{c}$ can be interpreted as defining the set of prices λ for the fixed resources (or assets) ($\mathbf{b}$) of the firm that would yield at least an equivalent return to the owner as producing a vector of products $\mathbf{x}$, which can be sold for prices ($\mathbf{c}$), from these resources. The dual constraint is a "greater than or equal to" because you can pay too much for a productive input, but market forces will ensure that you cannot count on buying an input for less than its value when turned into a saleable product.

Post multiplying the transpose of the technical input requirement matrix $\mathbf{A}$ by the dual prices λ results in an $m \times 1$ vector of marginal opportunity costs for each of the n potential production activities.

$\mathbf{A}'\lambda$ = vector of marginal opportunity costs of production.

For a single production activity x_i its opportunity cost of production for a vector of dual prices λ is:

$$\mathbf{a}_i' \lambda = (\text{column } i \text{ of } \mathbf{A})' (\lambda \text{ vector})$$

$$\mathbf{a}_i' \lambda = (a_{ij} \dots a_{mi}) \begin{pmatrix} \lambda \\ \vdots \\ \lambda_m \end{pmatrix} = \text{imputed cost of the last unit of } x_i \text{ produced, which equals the sum of imputed values of each resource needed times the amount of that resource needed to produce } x_i$$

$\uparrow$
 $\begin{pmatrix} \text{resource reqt.} \\ \text{(coefficients)} \\ \text{for input } x_i \\ \text{(one unit)} \end{pmatrix}$

$\cdot$

$\uparrow$
 $\begin{pmatrix} \text{Marg. imputed} \\ \text{value (for 1 unit)} \\ \text{of those resources} \\ \text{required} \end{pmatrix}$

$$\mathbf{a}_i' \lambda = \text{The cost of producing a unit of } x_i \text{ if I have to pay } \lambda \text{ for resources } \mathbf{b}$$

The Dual Constraint:

The constraint $A'\lambda \geq c$ can be interpreted for the production problem as saying that the Marginal Opportunity cost of producing a vector of x 's must be greater than or equal to the marginal revenue (c) for each of the x 's for a maximizing owner of the firm to sell at the firm value of $\lambda'b$.

Using the example of Rock Music production with an optimal solution that produces 18 units of Hiphop. The resources used in this production can be calculated from the coefficients in Table RR.

Hiphop production

Cost of Airplay time needed = $28 * 0.0$ + Cost of Studio time needed = $10 * 0.48 = 4.8$

$\$4.8 / \text{unit Hiphop} = c(\text{Hiphop}) = 4.8$.

Therefore Marginal opportunity cost = marginal revenue

Alternative production

Cost of Airplay time needed = $25 * 0.0$ + Cost of Studio time needed = $12 * 0.48 =$

$\$5.76 / \text{unit Alternative} > c(\text{Alternative}) = 3.5$.

Therefore Marginal opportunity cost > marginal revenue and Alternative CDs are not produced.

Showing the Primal/Dual Linkage

We want to show that the Primal optimality conditions imply that the dual constraints must hold.

Primal Problem

Given the following Primal problem,

$$(1) \quad \text{Maximize } c'x \quad \text{where } A = m \times n \quad \text{and } n > m$$

$$\text{s.t.} \quad Ax \leq b, \quad x \geq 0$$

Suppose there exists a basic feasible solution $x' = [x_B : 0]$ with a corresponding partition of A into the basis matrix B and non basis matrix D . It follows that:

$$(2) \quad x_B = B^{-1}b$$

The reduced cost vector r' equals:

$$(3) \quad r' = c'_D - c'_B B^{-1}D$$

If the solution to (1), x_B is optimal, for a maximization problem the scalar reduced cost is:

$$(4) \quad r_j = c'_D - c'_B B^{-1}d_j \leq 0 \quad \text{for all } j$$

Or equivalently stacking the d_j vectors together to yield the matrix D , the right side of (4) becomes:

$$(5) \quad c'_B B^{-1}D \geq c'_D$$

Defining the $1 \times m$ vector of dual variables:

$$(6) \quad \lambda' \equiv c'_B B^{-1}$$

Then by (5), $\lambda'D \geq c'_D$ at the optimum.

The shadow values of resources
you'd need to bring x_D 's into
basis
ie (The opportunity costs of
bringing activities into basis, times the
input requirement)

$\geq$

Marginal benefits of
bringing x_D 's into the
basis (Since $c_j x_j$ is the
benefit from x_j . $c_j x_j = 0$
when $x_j = 0$,(that is x_j is a
non-basic activity)

Dual Problem

The dual problem to (1) is specified as:

$$(7) \quad \begin{array}{ll} \text{Minimize} & \lambda' b \\ \text{Subject to} & A' \lambda \geq c \quad \text{and} \quad \lambda \geq 0 \end{array}$$

We want to show that the dual solution vector λ defined in (6) is:

(a) Feasible - That is, it satisfies the constraints in (7), and

(b) Optimal - By showing that the dual objective function value is equal to the optimal primal objective function value.

(A) Feasibility of λ

$$(8) \quad \lambda' A = [\lambda' B \quad \lambda' D] = [c'_B B^{-1} B \quad c'_B B^{-1} D] \\ = [c'_B \quad c'_B B^{-1} D]$$

but substituting the **inequality condition for optimality**

$c'_B B^{-1} D \geq c'_D$ in (5) we get:

$$(9) \quad [c'_B \quad c'_B B^{-1} D] \geq [c'_B \quad c'_D] = c'$$

combining (8) and (9):

$$(10) \quad \lambda' A \geq c' \quad \text{transposing yields} \quad A' \lambda \geq c$$

Thus by definition λ is a feasible solution to the constraints $A' \lambda \geq c$

(B) Optimality of λ

Using the dual objective function in (7) and our definition of λ in (6) we get:

$$(11) \quad \lambda' b = c'_B B^{-1} b = c'_B \underset{\uparrow}{x_B}$$

substitute the definition of $x_B = B^{-1}b$ from (2)

$\therefore$ Dual Objective Function value = Optimal Primal Objective Function value

$\therefore$ The dual value λ is optimal by the Strong Duality Theorem.*

The Point: **If the standard primal LP problem has an optimal basic feasible solution with a basis B then the vector $\lambda' \equiv c'_B B^{-1}$ is a feasible and optimal solution for the corresponding dual problem.**

***Strong Duality Theorem:** The theorem can be informally summarized as:

$\lambda' b = c'x$ If and only if x and λ are optimal solutions to the primal and dual problems respectively

Numerical Matrix Example - Yolo Model

The A matrix in Yolo is 4 x 4 . The * denotes the basic solution activities and binding constraints for the optimal solution.

		*		*	
		Alfalfa	Wheat	Corn	Tomato
A =	*Land	1.0	1.0	1.0	1.0
	Water	4.5	2.5	3.5	3.25
	Labor	6.0	4.2	5.6	14.0
	*Contract	0.0	0.0	0.0	33.25

The optimal solution to Yolo has two binding constraints Land and Contract, and two non-zero activities Wheat and Tomatoes. We collapse the A matrix to the basis matrix B **by removing the rows and columns that do not have * and do not constrain the optimal solution.** That is, if the rows are not binding, their coefficients are not in the basis.

The optimal basis matrix is therefore 2 x 2, $\therefore$ ignoring the slack Labor and Labor constraints. The optimal basis B is:

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 33.25 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$B^{-1} = \begin{bmatrix} 1 & -0.03007 \\ 0 & 0.03007 \end{bmatrix} \quad b = \begin{bmatrix} 600 \\ 6000 \end{bmatrix} \quad c_B = \begin{bmatrix} 160 \\ 825 \end{bmatrix}$$

(Note that for B to be invertible the number of activities in the basis **must equal** the number of binding constraints)

Check the matrix derivations against the computer printout.

(1) Optimal Primal Solution $x_B = B^{-1}b$ (forget about x_D ; they are all zeroes)

$$x_B = \begin{bmatrix} 1 & -0.03007 \\ 0 & 0.03007 \end{bmatrix} \begin{bmatrix} 600 \\ 6000 \end{bmatrix} = \begin{matrix} (1)(600) + (-0.03007)(6000) \\ (0)(600) + (0.03007)(6000) \end{matrix} = \begin{bmatrix} 419.58 \\ 180.42 \end{bmatrix}$$

(2) Optimal Dual Solution (Marginals on Resources)

$$\lambda' = c'_B B^{-1} = [160, 825] \begin{bmatrix} 1 & -0.03007 \\ 0 & 0.03007 \end{bmatrix}$$

λ values in Gams are the "MARGINALS" on resources. (They are > 0 for binding constraints, $= 0$ for non-binding constraints)

$$= [(160)(1) + (825)(0), (160)(-0.03007) + (825)(0.03007)]$$

$$= [160, 20]$$

(3) $r_j = c_j - z_j$ $z'_j = c'_B B^{-1}D = \lambda'D$

r_j values in Gams are the "MARGINALS" on activities.

(They $= 0$ on activities in basis, < 0 for non-basic activities)

$$z'_j = [160, 20] \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = [160, 160]$$

$$c_D = [121, 135]$$

$$\therefore r' = [121, 135] - [160, 160] = [-39, -25]$$

(4) Optimal Objective Function $\Pi = c'_B x_B$

$$\Pi = [160 \quad 825] \begin{bmatrix} 419.549 \\ 180.451 \end{bmatrix} = [67127.84 + 148872.07] = 216000$$

Parametric Analysis

Parametric analysis is the main method for obtaining policy results from optimization models. By changing the output prices or quantities of input available over a range of values we can generate empirical estimates of : (i) **The Supply Function** from the modeled process or (ii) **The Derived Demand** for inputs to the modeled process:

We may also be concerned with the **Sensitivity** of the model. Sensitivity tests the stability of conclusions from the model when we change a constraint or coefficient value. This is useful when we are uncertain of the exact value of a parameter and need to know whether knowing the value precisely is important.

Case: 1. Generating Derived Demand Functions for Inputs

The resource availability vector b is parameterized - that is, changed by small incremental values over a specified range. At each point the model is optimized and the value of the resource λ is plotted against the amount of the input b to form the derived demand function.

Given the Problem

$$\begin{array}{ll} \text{Max} & c'x \\ \text{subject to.} & Ax \leq b, \quad x \geq 0 \end{array}$$

The optimal solution is $x = [x_B \quad \vdots \quad 0]$

(Note: We assume the solution is not degenerate)

$$\therefore x_B = B^{-1}b$$

and $\Delta x_B = B^{-1}\Delta b$

Suppose Δx_B is small enough so there is no change in the basis, then we can define new x values, $\tilde{x}$ which are defined as:

$$\tilde{x} = [x_B + \Delta x_B \quad \vdots \quad 0] \quad \text{These result in a new objective function value:}$$

$$\begin{aligned} \text{Objective Function:} \quad & \text{old } \Pi = c_B' x_B \\ & \text{new } \tilde{\Pi} = c_B' (x_B + \Delta x_B) \\ & \Delta \Pi = \tilde{\Pi} - \Pi = c_B' \Delta x_B \\ & = c_B' B^{-1} \Delta b \\ & \Delta \Pi = \lambda \Delta b \end{aligned}$$

Selecting the b_i^{th} activity.

$$\therefore \frac{\Delta \Pi}{\Delta b_i} = \lambda_i \quad \text{Thus } \lambda \text{ measures the impact on the objective function of a marginal change in resource availability, given no change in the basis } B.$$

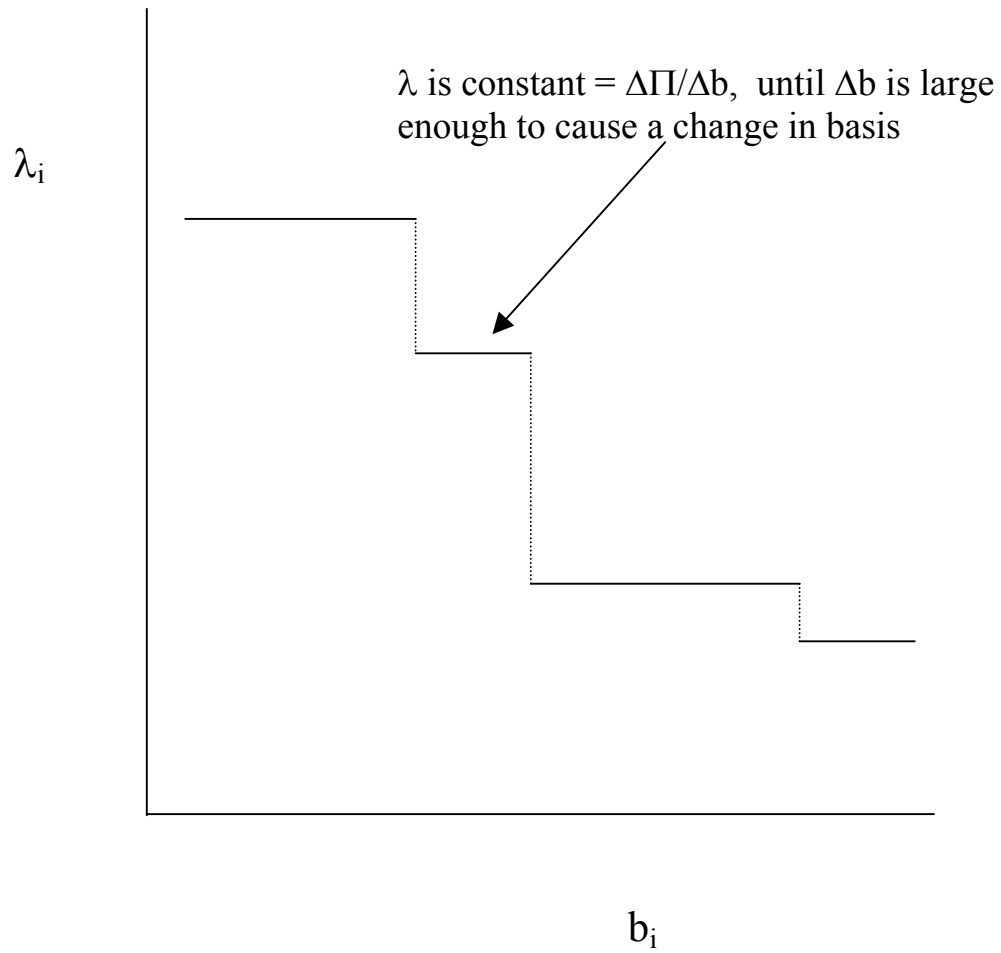
But from the equation for λ , ($\lambda' = c_B' B^{-1}$) we see that for linear problems, the value of

λ **does not change with Δb unless**

- (i) The basis B changes, or
- (ii) The objective function coefficients c_B are changed by parameterization.

Result

- I. For LP problems, the marginal change in the objective function for changes in resources ($\Delta \Pi / \Delta b$) is constant until the basis changes.
- II. When the basis changes, the value of λ changes due to new values in c_B' and B . This gives rise to the stepwise response to parameterization shown in the figure below.



Case: 2. Generating Supply Functions for Outputs

Similarly to the demand functions, supply functions are obtained by parameterizing a single objective function coefficient say, c_i , by small incremental values over a specified range. At each point the model is optimized and the quantity of the output x_i produced is plotted against the “price” to form the supply function. Clearly, for LPs there will only be a change in the product x_i when there is a change of basis caused by the change in c_i . Accordingly, the reaction when x_i is zero and therefore a non-basis activity is going to be different from the situation when $x_i > 0$, and is in the basis.

Case I Non-Basis activity supply parameterization

If $x_i = 0$, then from the optimality conditions we know that the reduced cost for this activity (assuming maximization)

$$r_i = c_i - z_i < 0$$

Since $z_i = \lambda' d_i$ and $\lambda' = c'_B B^{-1}$ the value of z_i is unchanged by a change in c_i because it is not in the vector c_B . However, if the value of c_i is increased by less than r_i then the reduced cost will still be < 0 (though reduced in numerical value) and x_i will still be a non-basis variable.

If $\Delta c_i > r_i$, then r_i will become positive and this will induce a change of basis which will bring the x_i activity into the basis with a positive value.

Case II Basis activity supply parameterization

Consider parameterizing a basis variable x_k . The coefficient c_k is now part of the vector c_B . Since x_k is an optimal basic variable we know that $r_k = c_k - z_k = 0$.

Again, $z_k = \lambda' d_k$ and $\lambda' = c'_B B^{-1}$ the value of z_k is therefore changed by a change in c_k and from the simplex criteria we know that:

- (i) The basis will change since Δc_k will cause $\Delta \lambda$, and some activities will leave the basis.
- (ii) The new basis will have a larger value for x_k since the reduced cost with the higher c_k value in the basis value vector will increase the value of r_k .

Thus whether we start with a non-basis or basic activity, parameterizing the objective coefficient over a discrete range will result in a series of stepped increases in the quantity of x_k in the optimal solution. The resulting supply function will be an upward sloping supply function.

Complementary Slackness

Primal Complementary Slackness

The concept of complementary slackness applies to both the primal and dual problems, but is easier to conceptually understand in the primal case. The formal proofs are developed for the dual CS case.

Given the standard LP problem of:

$$\begin{array}{ll} \text{Max} & c'x \\ \text{subject to.} & Ax \leq b, \quad x \geq 0 \end{array}$$

The primal complementary slackness theorem says that for the j th constraint:

$$\text{If } (b_i - a'_{ij} x) > 0 \quad \text{then } \lambda_i = 0$$

$$\text{Summarized as } (b_i - a'_{ij} x) \lambda_i = 0$$

Or in summation form

$$(b_i - \sum_j a_{ij} x_j) \lambda_i = 0$$

In words this says that if the total use of the i th input in all productive uses is less than the amount of input available, then the marginal scarcity value of additional units of input is zero.

Note there is a special case in which both $(b_i - \sum_j a_{ij} x_j)$ and λ_i are equal to zero. This says that the constraint is binding, but an additional unit of b_i will not add to the objective function value.

A corollary is:

$$\text{If } \lambda_i > 0 \text{ then } (b_i - \sum_j a_{ij} x_j) = 0.$$

That is to say, the shadow value cannot be positive if the constraint is not binding.

To get an intuitive idea of the primal complementary slackness, imagine that you are sunbathing on a large, sandy, uncrowded and hot beach. If you are offered additional sand for \$10 you are unlikely to purchase it, as your sand constraint is not binding. However if you are offered a cold drink (you do not have any) you are probably willing to pay a substantial premium price over the supermarket price to get a cold drink on the beach.

Dual Complementary Slackness

The Standard Dual Problem is:

$$\begin{aligned} \text{Min} \quad & \lambda' b \\ \text{s.t.} \quad & A' \lambda \geq c \quad \lambda > 0 \end{aligned}$$

Theorem.

Let x and λ be feasible solutions to the standard primal and dual problems. A necessary and sufficient condition for them to be optimal is:

$$(i) \quad \text{If } x_i > 0 \quad \text{then } \lambda' a_i = c_i$$

$$\text{Summarized by } [c' - \lambda' A] x = 0$$

$$(ii) \quad \text{If } \lambda' a_i > c_i \quad \text{then } x_i = 0.$$

The "Free Lunch" Theorem

Note x_i is never < 0 by definition of feasibility, and c_i is never $> \lambda' a_i$ by dual constraint that $\lambda' A \geq c$. The intuition for the direction of this constraint is that under economic (and optimizing) assumptions $\lambda' A < c$ cannot exist, as productive resources are never priced below their immediate productive value as this allows instantaneous capital gains. The possibility of such gains is ruled out by the "No Free Lunch" theorem.

Proof

Sufficiency (Sufficiency- If the conditions hold—then it is optimal.)

Assume condition (i) or condition (ii) holds. Then:

$$\begin{aligned} (1) \quad & [c' - \lambda' A] x = 0 \\ \therefore & \lambda' A x = c' x \quad \text{Since } A x = b \text{ for any basic feasible solution} \\ \therefore & \lambda' b = c' x \end{aligned}$$

$\therefore$ by the Duality theorem λ and x must be optimal if $\lambda' b = c' x$

Necessity (Necessity- If the problem is optimal—then the conditions hold).

Assume optimality - if λ^* and x^* are optimal solutions then from (1)

$$\begin{aligned} \lambda^*{}' b &= c^*{}' x^* && \text{dropping the } * \text{ notation} \\ \therefore \lambda^*{}' A x^* &= c^*{}' x^* \\ \therefore [c' - \lambda^*{}' A] x^* &= 0 \end{aligned}$$

Demonstrating Complementary Slackness

Case 1. Show that if $x_i > 0$, then $c_i' - \lambda'a_i = 0$

Verbally In the production problem "If something is produced in positive quantities ($x_i > 0$) then its marginal revenue equals its marginal opportunity cost at the optimum"

That is to say: $(c_i' - \lambda'a_i = 0)$

We select the subset of x , x_B (The basis Partition of x), since only the x_i 's > 0 are in the basis

Given that the solution is optimal it implies that $x_i > 0$ and all the r_i values $= 0$ (The " r_j " of each x_i is the net benefit of bringing x_i into the basis when x_i is already in the basis)

Writing the vector of r_i 's for the basis vectors B , then $r_i = (c_B' - c_B') = 0$.

Rewriting this

$$\begin{aligned} r_i &= c_B' - c_B' B^{-1}B \quad (\text{Since } r_j = (c_D' - c_B' B^{-1}D) \text{ when } x_j \text{ is not in the basis}) \\ \therefore c_B' - \lambda'B &= 0 \quad \text{since } r_i = 0 \text{ for basis activities} \\ \therefore c_B' &= \lambda'B \end{aligned}$$

(Note, the non-basis x_D are zero by definition $\therefore c' - \lambda'A$ becomes $c' - \lambda'B$)

Case 2. Show that if $x_i = 0$, then $c_i' - \lambda'a_i < 0$

Verbally In the production problem "If a product is not produced at the optimum, its net revenue must be less than its opportunity cost"

If $x = 0$ It implies that none of these x 's are in basis. Therefore we have the set x_D , the non-basis partition.

$$\therefore c' - \lambda'A \text{ becomes } c_D' - \lambda'D$$

$$\text{Recall } r_j = c_j - z_j = (c_D' - c_B' B^{-1}D) = (c_D' - \lambda'D)$$

$$\text{so the vector of } r_j\text{'s} = c_D' - \lambda'D$$

Since all the r_j 's < 0 for non-basic x_j 's at optimal solution

$$\therefore c_D' - \lambda'D < 0$$

$$\therefore \lambda'D > c_D' \quad \text{or} \quad c_j' - \lambda'd_j < 0$$

Note the strict inequalities

Corollary

If λ and x are optimal dual and primal solutions it implies that $[c' - \lambda' A] x = 0$ for all values.

Example: Auto Dealer Hype "Trust me, this car is selling below my invoice cost."

If you believe this—drop the course.

In this example, we are optimizing from the buyer's perspective. The buyer wants to minimize the price for a car with certain attributes. c_j is the dealer's invoice cost, and cars are traditionally sold as a vector of attributes and options on a base unit. Thus a "loaded" car will be composed of a base unit, air conditioning, FM stereo, "dealer's prep", etc. These attributes of the car are represented by the $m \times 1$ vector a_j .

The dealer's problem is to assign a set of prices λ to the $m \times 1$ a_j vector such that

$$\lambda a_j - c_j \geq 0$$

Remember that prices on a car lot are always negotiable, and the buyer wants to minimize the cost "out of the door" of a car with one of each of the $m \times 1$ set of attributes. That is to minimize $\lambda' b$. The dealer wants to convince you that at the price vector λ the complementary slackness theorem does not hold. The complementary slackness theorem says that if the dealer truly has set prices λ below his invoice then:

$$\lambda a_j - c_j < 0$$

and the dealer will set $x_j = 0$. That is the car which is priced below invoice is never in stock when you get to the dealer since "It has just been sold."

Duality and Reduced Cost

Since λ is the vector of opportunity costs on the binding resources, and r_j 's form a vector of revenue minus activity opportunity costs, they must be connected.

$$\lambda = c_B B^{-1}$$

$$r = [c_D - c_B B^{-1} D]$$

$$= c_D - \lambda D$$

Each individual element in the vector r , $r_j = c_j - \lambda d_j$

$$\text{where } \lambda d_j = [\lambda_1 \dots \lambda_m]$$

$$\begin{array}{c} \lambda_1 \\ \vdots \\ \lambda_m \end{array} \begin{array}{c} d_{1j} \\ \vdots \\ d_{mj} \end{array}$$

Therefore λd_j is equal to the shadow value of inputs required to produce a unit of activity x_j that is not currently in the basis.

V CALIBRATING OPTIMIZATION MODELS

Even with a constraint structure and parameters that are theoretically correct, it is highly unlikely that a model will calibrate closely to the base year data. This is inherent in the structure of models which are by definition simplified abstractions of the real systems. In the process of abstracting and simplifying a real system the model loses information and needs to be verified against actual behavior. Just as in econometric modeling there are two phases of model estimation and prediction, in optimization modeling there are the two phases of model calibration and simulation.

Fundamentally, the calibration process is one of using a hypothesized function and data on input and output levels in the base year to derive specific model parameter values that "close" the model. By closing the model, we mean that the calibration parameters lead to the objective function being optimized for the base year conditions. Like econometrics, calibration methods assume that the observed actions are motivated by optimizing over some set of criteria. It follows that if we can derive the parameters, that when optimized, lead to the observed actions, we have derived the parameters most likely to have been used by the decision maker.

Given that most optimization models are specified to optimize profits to economic decision makers, they can in theory be calibrated from the demand (price) or supply (cost) sides. Calibrating model prices by deriving demand parameters has been practiced by builders of large quadratic programming models for many years. However demand side calibration can only help when the model is on a large enough scale so that changes in output change product prices. In addition, because models are usually specified with several regions supplying a single market demand, using a small number of demand parameters cannot calibrate the cropping patterns over several regions.

Calibrating on the Model Supply Side

This section contains a short overview and critique of the traditional methods of calibrating the supply side of optimization models using linear constraints. The shortcomings of the constraint calibration methods lead to a discussion of ways to derive nonlinear supply functions that are based on observed behavior by decision makers, but calibrate the model in general. These methods are termed "Positive Mathematical Programming" (PMP) (Howitt 1995) since they are based on positive inferences from the base year data, rather than normative assumptions.

A Review of Supply Side Constraint Calibration

Programming models should calibrate against a base year or an average over several years. Policy analysis based on normative models that shows a wide divergence between base period model outcomes and actual production patterns is generally unacceptable. However, models that are tightly constrained can only produce the subset of normative results that the calibration constraints dictate. The policy conclusions are thus bounded by a set of constraints that are expedient for the base year but often inappropriate under policy changes. This problem is exacerbated when the model is built on a regional basis with very few empirical constraints but a wide diversity of crop production. For example, the Yolo model presented in the previous chapter is highly simplified with only four cropping activities, but still requires unrealistic constraints on the amount of labor that can be employed to produce all four crops in the optimal solution. While labor requirements do vary, they are subject to a labor supply. The supply function will increase labor available in a given quarter at an increased cost of overtime or operations

performed by custom operators. To suggest that a more profitable crop in some policy scenario will never be grown beyond a certain limit because of labor constraints is to radically depart from the actual empirical solution. In addition to the labor constraint, the water constraint may also be artificially constraining if the farmers have access to groundwater or river pumping. More complex linear production models require more complicated constraint structure to reproduce the observed cropping pattern. In many traditional models the proportions of crops are restricted by "rotational" constraints or "flexibility" constraints. These constraints determine the optimal solution not only for the base year for which they are appropriate, but also for policy runs that attempt to predict the outcome of changed prices, costs or resource availability. The solution of the model under policy runs is therefore significantly restricted by the base year solution constraints.

This section is a brief overview of some of the past calibration methods in mathematical programming models. For a more comprehensive discussion see Hazell and Norton (1986) or Bauer and Kasnakoglu (1990). It is worth noting that no single linear constraint calibration method has proved sufficiently satisfactory to dominate the mathematical programming literature.

Previous researchers such as Day (1961) have attempted to provide added realism by imposing upper and lower bounds to production levels as constraints. McCarl (1982) advocated a decomposition methodology to reconcile sectoral equilibria and farm level plans. Both approaches require additional micro level data and result in calibration constraints influencing policy response.

Meister, Chen, and Heady (1978), in their national quadratic programming model, specify 103 producing regions and aggregate the results to 10 market regions. Despite this structure, they note the problem of overspecialization and suggest the use of rotational constraints to curtail the overspecialization. However, it is comparatively rare that agronomic practices are fixed at the margin, more commonly they reflect net revenue maximizing trade-offs between yields, costs of production, and externalities between crops. In this latter case, the rotations are themselves a function of relative resource scarcity, output prices, and input costs.

Hazell and Norton (1986) suggest six tests to validate a sectoral model. First, a capacity test checks whether the model constraint set allows the base year production. Second, a marginal cost test ensures that the marginal costs of production, including the implicit opportunity costs of fixed inputs, are equal to the output price. Third, compare the dual value on land with actual rental values. Three additional comparisons of input use, production level, and product price are also advocated. Hazell and Norton show that the percentage of absolute deviation for production and acreage over five sectoral models ranges from 7 percent to 14 percent deviation. The constraint structures needed for this validation are not defined.

In contrast, the PMP approach (Howitt 1995) achieves exact calibration in acreage, production and price. The PMP approach was applied to the Turkish Agricultural Sectoral Model (TASM) which is one of the models listed by Hazel and Norton. The resulting PMP version of TASM calibrated exactly with the base year (Bauer and Kasnakoglu 1988) and showed consistency in the parameters over the seven years used for calibration. A recent application of a PMP calibrated model (U S Bureau of Reclamation, 1997) has been built to analyze the effects of large inter-sectoral water reallocations in California. The model termed the Central Valley Production Model (CVPM) was tested by out of sample predictions of regional crop acreage changes during a recent drought period. The CVPM . The predictions were close with three contract crops (Sugar beet, Tomatoes, and Subtropical orchard)

having a 14- 23 % error. The remaining nine crops had prediction errors below 7%. Regional crop acreage was predicted for eleven regions. For all of the regions the crop acreage predictions had errors below six percent.

The calibration problem for farm level, regional, and sectoral LP models can be mathematically defined by the common situation in which the number of binding constraints in the optimal solution (m) are less than the number of non-zero activities (n) observed in the base solution. If the modeler is fortunate enough to have empirical data to specify, *a priori*, a constraint set that reproduces the optimal base year solution, then additional model calibration is clearly redundant. The PMP approach is developed for the majority of empirical model builders who, for lack of empirical justification, data availability, or cost, find that the empirical constraint set does not reproduce the base year result. The LP solution is an extreme point of the binding constraints. In contrast, the PMP approach views the optimal farm production as a boundary point which is a combination of binding constraints and first order conditions.

Relevant constraints should be based on either economic logic or the technical environment under which the agricultural production is operating. Calibration problems are especially prevalent where the constraints represent allocatable inputs, actual rotational limits and policy constraints. When the basis matrix has a rank less than the number of observed base year activities, the resulting optimal solution will suffer from overspecialization bias of production activities compared to the base year.

A root cause of these problems is that linear programming was originally used as a normative farm planning method where full knowledge of the production technology is assumed. Under these conditions, any production technology can be represented as a linear Leontief specification, subject to resource and stepwise constraints. For aggregate policy models, this normative approach over simplifies the production and cost technology due to inadequate knowledge. In most cases, the only regional production data is an average or "representative" figure for crop yields and inputs. This common data situation means that the analyst using linear production technology in programming models is attempting to estimate marginal behavioral reactions to policy changes, based on average data observations. The average conditions can be assumed to be equal to the marginal conditions only where the policy range is small enough to admit linear technology over the whole range.

Two broad approaches have been used to reduce the specialization errors in optimizing models. The demand-based methods have used a range of methods to add risk or endogenize prices. These have reduced the calibration, problem but substantial calibration problems remain in many models, Just (1993).

A common alternative approach is to constrain the crop supply activities by rotational (or flexibility) constraints or step functions over multiple activities (Meister, Chen, and Heady, 1978). In regional and sectoral models of farm production, the number of empirically justifiable constraints are comparatively few. Land area and soil type are clearly constraints, as is water in some irrigated regions. Crop contracts and quotas, breeding stock, and perennial crops are others. However, it is rare that some other traditional programming constraints such as labor, machinery, or crop rotations are truly restricting to short-run marginal production decisions. These inputs are limiting, but only in the sense that once the normal availability is exceeded, the cost per unit output increases due to overtime, increased probability of machinery failure or disease. In this situation the analyst has a choice. If the assumption of linear production (cost) technology is retained, the observed output levels infer that additional binding

constraints on the optimal solution should be specified. Comprehensive rotational constraints are a common example of this approach.

An alternative explanation of the situation where there are more crop activities than constraints is that the profit function is nonlinear in land for most crops, and the observed crop allocations are a result of a mix of unconstrained and constrained optima. The equilibrium conditions for this case would be satisfied if some of all of the cropping activities had decreasing returns to land as acreage was increased. The most common reasons for a decreasing gross margin per acre are declining yields due to heterogeneous land quality, risk aversion, or increasing costs due to restricted management or machinery capacity.

Positive Mathematical Programming

The positive mathematical programming (PMP) approach is being adopted quite rapidly for agricultural sector models. In their introduction to their book on “Agricultural Sector Modelling and Policy Information Systems”(2001) Heckelee et al state that they received

“a rich supply of about 60 proposals from 16 countries reflecting the breadth of work directed to agricultural sector modeling and policy information systems. Because the sample of proposals is indicative of current emphasis in research, it is worth mentioning that, from a methodological point of view almost 25% of the supply might be called “econometric partial” analyses, another 25% “programming models” (half of which relying on PMP) whereas the remaining half of the proposals covered a multitude of quantitative methods..”

Behavioral Calibration Theory

The process of calibrating models to observed outcomes is an integral part of constructing physical and engineering models but is rarely formally analyzed for optimization models in agricultural economics. In this section we show that observed behavioral reactions yield a basis for calibrating models in a formal manner that is consistent with microeconomic theory. Analogously to econometrics, the calibration approach draws a distinction between the two modeling phases of calibration (estimation) and policy prediction.

On a regional level, the information on the product output levels and farm land allocations is usually more accurate than the estimates of marginal crop production costs. This is particularly true when micro data on land class variability, technology, and risk feature in the farmers' decisions, but are absent in the aggregate cost data available to the model builder. Accordingly, the PMP approach uses the observed acreage allocations and outputs to infer marginal cost conditions for each regional crop allocation observed. This inference is based on parameters that are known to be accurately observed and the usual profit maximizing and concavity assumptions.

The Nonlinear Calibration Proposition states: If the model does not calibrate to observed production activities with the full set of general linear constraints that can be empirically justified. A necessary condition for profit maximization at the observed values is that the objective function is nonlinear in at least some of the activities.

Many regional models have some nonlinear terms in the objective function reflecting endogenous price formation or risk specifications. Although it is well known that the addition of nonlinear terms improves the diversity of the optimal solution, there is usually an insufficient number of independent nonlinear terms to accurately calibrate the model.

The Calibration Dimension Proposition. The ability to calibrate the model with complete accuracy depends on the number of nonlinear terms that can be independently calibrated.

The ability to adjust some nonlinear parameters in the objective function, typically the risk aversion coefficient, can improve model calibration. However, if there are insufficient independent nonlinear terms the model cannot be made to calibrate precisely. In technical terms, the number of instruments the modeler has available to calibrate the model may not span the set of activities that need to be calibrated. For proofs of these latter two propositions, see the appendix of Howitt (1995)

A Simple Cost Based Approach to PMP Calibration.

This section demonstrates the PMP calibration process using nonlinear costs and constant yields to calibrate the model. The derivation is shown in its simplest form. Once you have the concept, the more complex development of changing yields will be clearer.

The key concept of PMP calibration developed in work by Fiacco & McCormack is that every linear constraint in an optimization problem can also be modeled by a nonlinear cost function with appropriately chosen coefficients.

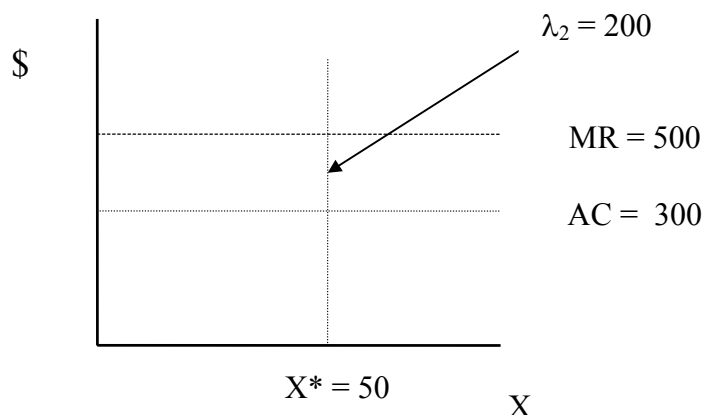
An Initial Primer on Cost Based PMP Calibration – Single Crop

A single linear crop production activity is measured by the acres x allocated to the crop. The yield is assumed constant. The data available to the modeler is:

Marginal revenue / acre is assumed constant at	\$500 / acre
Average Cost is	\$300/ acre
Observed acres allocated to the crop	50 acres

In the first step a measure of the value of the residual cost needed to calibrate the crop acreage to 50, by setting marginal revenues equal to marginal cost at that acreage is obtained from a constrained linear program.

Figure 1.
LP Constrained by calibration constraints



1. From the nonlinear calibration proposition we know that either (or both) the cost or production function is nonlinear if we need calibration constraints. In this case we define the total cost function to be quadratic in acres (x). There are very many possible nonlinear forms, but this is the simplest general form.

$$TC = \alpha x + \frac{1}{2} \gamma x^2$$

2. Under optimization, if unconstrained, crop acreage expands until the marginal cost equals marginal revenue. Therefore $MC = MR$ at $x = 50$

3. It follows that the value λ_2 in the linear model is the difference at the constrained calibration value and is equal to $MR - AC$. But from step 2 we know that $MR = MC$ therefore $\lambda_2 = MC - AC$, since $MR = MC$ at $x = 50$

Given the hypothesized total cost function TC :

$$MC = \alpha + \gamma x$$

$$AC = \alpha + \frac{1}{2} \gamma x$$

$$\text{therefore } MC - AC = \alpha + \gamma x - (\alpha + \frac{1}{2} \gamma x)$$

$$\text{Therefore } \lambda_2 = MC - AC = \frac{1}{2} \gamma x$$

and the cost slope coefficient is calculated as:

$$\gamma = 2\lambda / x^* = (2 * 200) / 50 = 8$$

4. We can now calculate the value of α using the AC information in the basic data set as:

$$300 = \alpha + \frac{1}{2} * 8 * 50$$

$$\text{Therefore } \alpha = 300 - 200 = 100$$

5. Using the values for α and γ the unconstrained quadratic cost problem is:

$$\begin{aligned} \text{Max } \Pi &= 500x - \alpha x - \frac{1}{2} \gamma x^2 \\ &= 500x - 100x - \frac{1}{2} 8 x^2 \end{aligned}$$

$$\delta \Pi / \delta x = 500 - 100 - 8x$$

Solving for $\delta \Pi / \delta x = 0$ which implies $MR = MC$

$$8x = 400 \quad \text{therefore } x^* = 50.$$

NOTE that the unconstrained model calibrates exactly in x and also in Π

Also note that:

- (a) $MC = MR$ at $x = 50$
- (b) $AC = 300$ at $x = 50$
- (c) The cost function has been “tilted”
- (d) Two types of information are used x^* and AC

- (e) The observed x^* quantities need to be mapped into value space by the LP before it can be used
- (f) The model now reflects the preferences of the decision maker
- (g) The model is unconstrained by calibration constraints for policy analysis

An Analytic Derivation of Calibration Functions

This section will show that the PMP non-linear calibration approach can be applied to any non-degenerate linear problem. The derivation of the general result proceeds in three steps. The first step shows that the dual value on the calibration constraint for the calibrated activity set x_k is equal to the reduced cost of the activity x_i in the un-calibrated base problem. The second step shows that if the correct non-linear penalty function is added to the objective function, the resulting nonlinear problem satisfies the necessary conditions for optimality at the required value of x_i . Finally, it is shown that the correct penalty function has a gradient at the required value of x_i equal to the negative of the calibration dual

The general linear programming optimization problem can be compactly written as:

$$(1) \quad \begin{array}{ll} \text{Max}_x & c'x \\ \text{Subject to} & Ax \leq b \\ & x \geq 0 \end{array}$$

Where c is an $p \times 1$ vector of net returns per unit activity, and the matrix A and right hand side b are the usual technical constraints and right hand sides. The dimension of A is $m \times p$, ($m < p$). The basis dimension of A is m . If the number of observed activities is n ($n \leq p$), where $n = k + m$, then in addition to the x_m basis activities, there are an additional k activities x_k that are observed and need to be calibrated into the optimal model solution. For simplicity, define the LP problem as only subject to one set of upper bound calibration constraints:

$$(2) \quad \begin{array}{ll} \text{Max}_x & c'x \\ \text{Subject to} & Ax \leq b \\ & Ix \leq \tilde{x} + \varepsilon \\ & x \geq 0 \end{array}$$

where the ε is added to the calibration constraints to prevent degeneracy.

The optimal basic solution to this problem will have a mix of n binding resource and calibration constraints. The A matrix can be partitioned into an $m \times m$ basis matrix B that corresponds to the m least profitable "marginal" activities (x_m), and an associated $m \times k$ matrix N for the k calibrated

activities x_k . Dropping out the $p - n$ zero activities the optimal basic solution to problem (2) can be written as:

$$(3) \quad \begin{aligned} & \text{Max}_x \quad c'_m x_m + c'_k x_k \\ & \text{subject to} \quad \hat{A}\hat{x} = \hat{b} \\ & \text{or partitioned as} \\ & \begin{bmatrix} B & N \\ 0 & I \end{bmatrix} \begin{bmatrix} x_m \\ x_k \end{bmatrix} = \begin{bmatrix} b \\ \tilde{x}_k + \varepsilon \end{bmatrix} \end{aligned}$$

The optimal dual constraints for this problem are:

$$(4) \quad \begin{aligned} & \hat{A}'\lambda = c \\ & \text{which is partitioned as} \\ & \begin{bmatrix} B' & 0 \\ N' & I \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} c_m \\ c_k \end{bmatrix} \end{aligned}$$

Equation (4) can be solved for the values of λ_1 and λ_2 by inverting the partitioned constraint matrix, using the partitioned form of the inverse, to yield:

$$(5) \quad \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} P & 0 \\ Q & I \end{bmatrix} \begin{bmatrix} c_m \\ c_k \end{bmatrix}$$

$$\text{Where } P = B'^{-1} \quad \text{and} \quad Q = -N' B'^{-1}$$

From equation (5) we see that the $k \times 1$ vector of dual values for the binding calibration constraints has the value:

$$(6) \quad \lambda_2 = c_k - N' B'^{-1} c_m$$

The right hand side of equation (6) is difference between the gross margin of the calibrating activity c_k and the equivalent gross margin that can be obtained from the less profitable marginal cropping activities c_m . In other words λ_2 is the marginal opportunity cost of restricting the calibrated activities x_k by the amount needed to bring the marginal x_m activities into the expanded basis. This cost of restricting the more profitable activities x_k in the basis is similar to the familiar reduced cost term.

Note: First that when land is the numeraire, the corresponding coefficients in the N and B matrices are one.

Second, that the sign on λ_2 is positive for Gams printouts as a marginal increase in the right hand side upper bound on the more profitable activities will increase the objective function value.

The dual values associated with the set of binding calibration constraints (λ_2) are independent of the resource and technology constraint dual values (λ_1), since the constraint decoupling proposition shows that the values for λ_1 are not changed by the addition of the calibration constraints

If an increasing nonlinear cost function is added to the objective function for the x_K activities that need to be calibrated the marginal and average costs of producing x_K will differ. The net return to land from x_K now decreases as the acreage is increased. The net returns to land from x_K reach an internal equilibrium solution at the point where they are equal to the opportunity cost of land set by the marginal crops x_M .

If the calibration constraints are removed and a nonlinear cost for x_K is added, problem (2) becomes:

$$(7) \quad \begin{aligned} \text{Max}_{x_k} \quad & c'_m x_m + c'_k x_k - f(x_k) \\ \text{Subject to} \quad & [B \quad N] \begin{bmatrix} x_m \\ x_k \end{bmatrix} = b \end{aligned}$$

The reduced gradient (the nonlinear equivalent of the reduced cost) for activities x_k is derived by rewriting the set of binding constraints in (7) so that x_m is a function of x_k :

$$(8) \quad x_m = B^{-1}b - B^{-1}N x_k$$

Substituting the expression for x_m back into the objective function in (7) defines the problem as an unconstrained function of x_k .

$$(9) \quad \text{Max} \quad c'_m (B^{-1}b - B^{-1}N x_k) + c'_k x_k - f(x_k)$$

The unconstrained gradient of this nonlinear problem in x_k is defined as the reduced gradient and is:

$$(10) \quad \begin{aligned} & c_k - N' B'^{-1} c_m - \nabla f'(x_k) \\ \text{where } & \nabla f(x_k) \text{ is the gradient of } f(x_k) \end{aligned}$$

Luenberger (1984) shows that a zero valued reduced gradient is a necessary condition for the optimum of a nonlinear problem. The calibrated problem (7) will optimize with a zero reduced gradient at the values $\tilde{x}_k$ when $c_k - N' B'^{-1} c_m = \nabla f'(x_k)$ or substituting into equation (6), when $\nabla f'(x_k) = \lambda_2$.

Proposition:

If the parameters of $f(x)$ are calibrated such that at the value $\tilde{x}_k$,

$$\nabla f'(\tilde{x}_k) = \lambda_2$$

the model will be optimal exactly at the calibrating acres.

To reiterate, equation (6) shows that λ_2 is equal to the first two terms of equation (10). It follows that the reduced gradient of the resulting nonlinear problem will equal zero at $\tilde{x}_k$. As this is a necessary condition for the optimum, the problem in equation (7) will calibrate at the values $\tilde{x}$ without calibration constraints.

Equation (9) shows that optimal solution to the calibrated problem responds to changes in the linear gross margin “c”, the right hand side values “b”, or the constraint coefficients “B or N”.

The economic interpretation of the calibration cost function $f(x_K)$ is as follows. From equation (6) we see that λ_2 is equal to the difference between the gross margins per unit land for x_K and x_M . The gross margins are calculated from the observed average variable costs. It follows that:

$$(11) \quad \lambda_2 = (P_K Y_K - AC_K) - (P_M Y_M - AC_M)$$

but since the gross margin for the marginal crop x_M is equal to the opportunity cost of land, and since the land coefficients in N and B equal one, using equations (5) and (6) we can rewrite (11) as:

$$(12) \quad \lambda_2 = (P_K Y_K - AC_K) - \lambda_1 \quad \text{or} \quad \lambda_2 + \lambda_1 = (P_K Y_K - AC_K)$$

but by the equimarginal principle, at the optimal allocation of land, all crops must have a marginal net return equal to the opportunity cost of land, therefore at the optimal solution to the nonlinear problem defined by equation (7):

$$(13) \quad \lambda_1 = P_K Y_K - MC_K = P_K Y_K - AC_K - (MC_K - AC_K)$$

substituting (12) into (13) yields:

$$(14) \quad \lambda_2 = MC_K - AC_K$$

To summarize, this section has shown that linear and nonlinear optimization problems can be exactly calibrated by the addition of a specific number of nonlinear terms. We have used a general nonlinear specification to show that since the calibrated problem (2) yields the necessary conditions (6). If the nonlinear problem (7) has a nonlinear cost function $f(x_K)$ that satisfies equations (6), (10), and (14), the resulting nonlinear problem will calibrate exactly in the primal and dual values of the original problem, but without any inequality calibration constraints.

In the next section we show how the calibration procedure can be simply implemented using a quadratic cost function in a two-stage process that is initiated with a calibrated linear program.

An Empirical Calibration Method

The previous section showed that if the correct nonlinear parameters are calculated for the (k- m) unconstrained (independent) activities, the model will exactly calibrate to the base year values x without additional constraints. The problem addressed in this section is to show how the calibrating parameters can be simply and automatically calculated using the minimal data set for a base year LP.

Given that nonlinear terms in the supply side of the profit function are needed to calibrate a production model, the task is to define the simplest specification that is consistent with the technological basis of agriculture, microeconomic theory and the data base available to the modeler.

A highly probable source of nonlinearity in the profit function is due to heterogeneous land quality. This will cause the marginal cost per unit of output to increase as the proportion of a crop in a specific area is increased. This phenomenon, first formalized by Ricardo (Peach (1993)), is widely noted by farmers, agronomists, and soil scientists, but often omitted from quantitative production models.

Defining yields per acre as constant and marginal cost as increasing in land allocation is a considerable simplification of the complete production process. Given the applied goal of this "positive" modeling method, the calibration criteria used is not whether the simple production specification is true, but rather, does it capture the essential behavioral response of farmers, and can it be made to work with the restricted data bases and model structures available.

The output q_i from a given cropping activity "i" under a Leontieff production specification with land x_i and two other inputs is :

$$(15) \quad q_i = \text{Min} (x_i, a_{i2}x_i, a_{i3}x_i) y_i$$

In many LP problems such as the Yolo example, the objective function coefficients c_i represent the Gross margin per acre faced by the farmer. The c_i coefficient is composed of the product of the average yield, the price per unit output, with the average variable costs deducted. Since we are specifying the yield as constant but the marginal cost as increasing with acreage in this model, we have to separate these components.

The calibrated optimization problem equivalent to equation (7) with land as the restricting factor becomes

$$(16) \quad \text{Max } \sum_i P_i Y_i x_i - (\alpha_i + 0.5 \gamma_i x_i)x_i - \sum_{j=2}^3 \omega_j a_{ij} x_i$$

$$\text{subject to } Ax \leq b \quad \text{and } x \geq 0$$

where $a_{i1} = 1$ and $A = (m \times n)$ with elements a_{ij} , x_i is the acreage of land allocated to crop i, Y_i is the yield per acre, α_i and γ_i are respectively the intercept and slope of the marginal cost function per acre, and ω_j is the cost per unit of the j^{th} input.

The PMP calibration approach uses three stages. In the first stage a constrained LP model is used to the dual values for both the resource and calibration constraints, λ_1 and λ_2 respectively. In the second stage, the calibrating constraint dual values (λ_2) are used, along with the data based average yield function, to uniquely derive the calibrating cost function parameters (α_i and γ_i). In the third stage, the cost parameters are used with the base year data to specify the PMP model in equation (16). The resulting model calibrates exactly to the base year solution and original constraint structure.

The procedure is illustrated using a very simple problem which has a single land constraint (5 acres) and two crops wheat and oats. The following parameters are used:

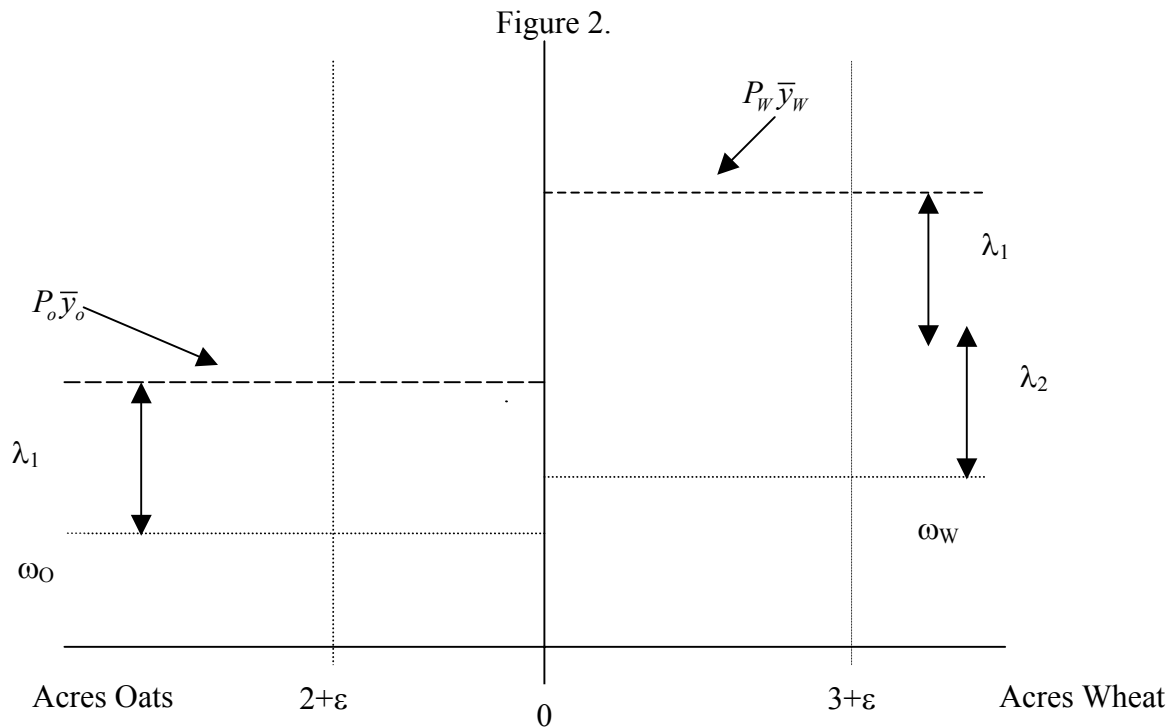
Wheat (w)

(Oats) (o)

Crop prices	$P_W = \$2.98/\text{bu.}$	$P_O = \$2.20/\text{bu.}$
Variable cost/acre	$\omega_W = \$129.62$	$\omega_O = \$109.98$
Average yield/acre	$\bar{y}_W = 69 \text{ bu.}$	$\bar{y}_O = 65.9 \text{ bu.}$

The observed acreage allocation in the base year is 3 acres of wheat and 2 acres of oats.

Figure 2 shows the initial problem in a diagrammatic form for two activities, with one resource constraint and two upper bound calibration constraints. Note that at the optimum, the calibration constraint will be binding for wheat, the activity with the higher average gross margin, while the resource constraint will restrict the acreage of oats.



The problem in Figure 2 is:

$$\begin{aligned}
 (17) \quad & \text{Max } (2.98*69 - 130) x_W + (2.20*65.9 - 110) x_O \\
 & \text{subject to} \quad (i) \quad x_W + x_O \leq 5 \\
 & \quad \quad \quad (ii) \quad x_W \leq 3.01 \\
 & \quad \quad \quad (iii) \quad x_O \leq 2.01
 \end{aligned}$$

Note the addition of the ε perturbation term (0.01) on the right hand side of the calibration constraints. The average gross margin from wheat is \$76/acre and oats \$35/acre. The optimal solution to the stage 1 problem is when the wheat calibration constraint is binding at a value of 3.01 and constraint (i) is binding when the oat acreage equals 1.99. The oat calibration constraint is slack.

Two equations are solved for the two unknown yield parameters (α and γ). Using the quadratic total land cost function specified in (16) and the first order conditions in equation (10), the first equation for the marginal cost coefficient sets λ_2 equal to the difference between marginal and average cost based on equation(14) and derives the calibrated value for γ .

$$\begin{aligned}
 (18) \quad & \nabla f'(\tilde{x}_k) = \lambda_{2k} \\
 & 0.5 \gamma_k \tilde{x}_k = \lambda_{2k} \\
 & \gamma_k = \frac{2\lambda_{2k}}{\tilde{x}_k}
 \end{aligned}$$

Equation (18) uses the value of the dual on the LP calibration constraint (λ_2) which is shown in figure 1 to be the difference between the average cost (AC) of the crop and the marginal cost (MC).

The second equation is the average cost for crop i, $\bar{y}_i$:

$$\begin{aligned}
 (19) \quad & \sum_{ij} \omega_{ij} a_{ij} = c_i = \alpha_i + 0.5 \gamma_i x_i \\
 & \therefore \alpha_i = c_i - 0.5 \gamma_i x_i
 \end{aligned}$$

The derivation of the two types of dual value λ_1 and λ_2 , can be shown for the general case (Howitt 1995). The A matrix in (2) is partitioned by the optimal solution into an $m \times m$ matrix B associated with the marginal variables x_m , an $m \times 1$ subset of x with inactive calibration constraints. The second partition of A is into an $m \times k$ matrix N associated with a $k \times 1$ partition of x, x_N of non-zero activities constrained by the calibration constraints. The equation for λ_1 is the usual LP form of :

$$(20) \quad \lambda_1 = c'_m B^{-1}$$

The elements of vector x_m are the acreages produced in the crop group limited by the general constraints, and λ_1 are the dual values associated with the set of $m \times 1$ binding general constraints. Equation (20) states that the value of marginal product of the constraining resources is a function of the revenues from the constrained crops. The “independent” crops (x_k) do not influence the dual value of the resource by the decoupling proposition. This is consistent with the principle of opportunity cost in which the marginal net return from a unit increase in the constrained resource determines its opportunity cost. Since generally the more profitable crops x_k are constrained by the calibration constraints, the less profitable crop group x_m are those that could use the increased resources and hence set the opportunity cost.

Equation (21) determines the dual values on the upper bound calibration constraints on the crops.

$$(21) \quad \lambda_2 = -N' B'^{-1} c_m + I c_k \text{ and substituting (20)}$$

$$= I c_k - N' \lambda_1$$

The dual values for the binding calibration constraints are equal to the difference between the marginal revenues for the calibrated crops (x_k) and the marginal opportunity cost of resources used in production of the constrained marginal crops (x_m). Since the stage I problem in Figure 2 has a linear objective function, the first term in (21) is the crop average value product of land in activities x_k . The second term in (21) is the marginal opportunity cost of land from equation (20). In this PMP specification, the difference between the average and marginal cost of land is attributed to changing land quality. Thus the PMP dual value (λ_2) is a hedonic measure of the difference between the average and marginal cost of land, for the calibrated crops. Analogously to revealed preference, PMP can be thought of as revealed efficiency based on observed land allocations.

Equation (21) substantiates the dual values shown in Figure 2, where the duals for the calibration constraint set (λ_2) in the stage I problem are equal to the divergence between the LP average cost per acre and the marginal opportunity cost per acre.

The dual value on land (λ_1) is \$35 and on the two calibration constraints (λ_2) = [41 and 0]. Using equation (14), the λ_2 value for wheat and the base year data, the cost function slope for wheat is calculated as:

$$(22) \quad \gamma_w = 2 * 41 / 3 = 27.333$$

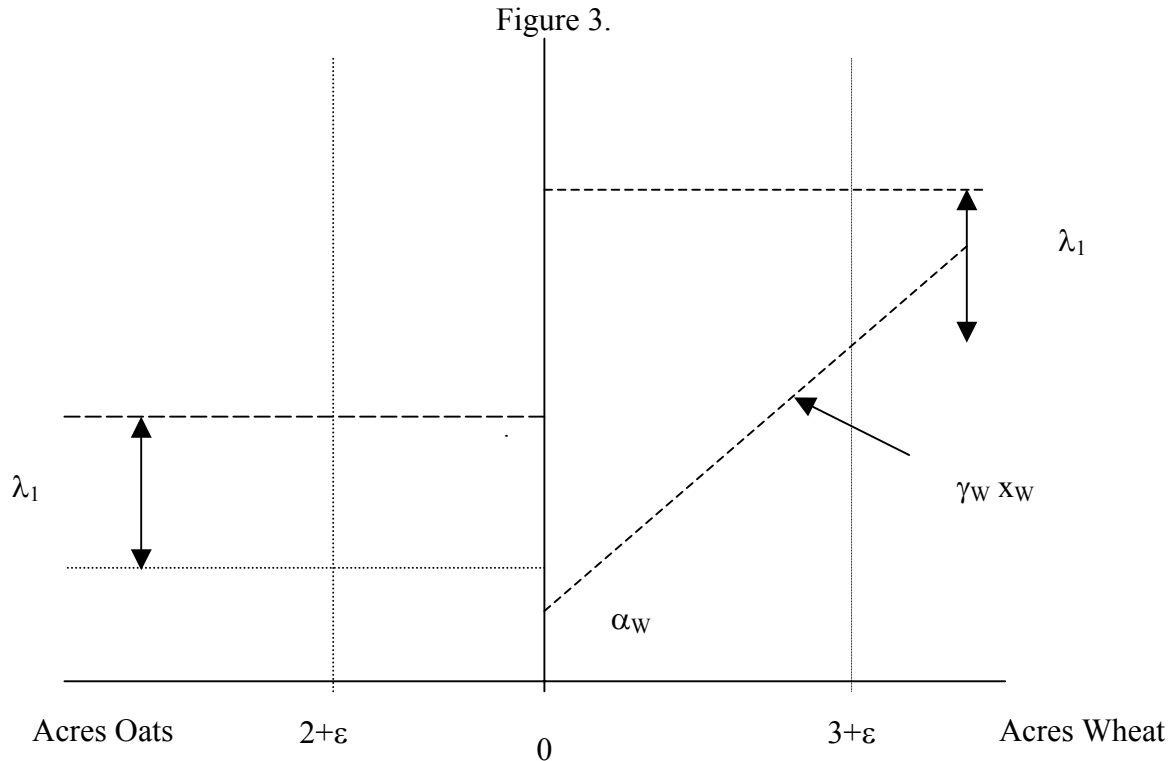
γ_w is now substituted into equation (19) to calculate the cost intercept α_w .

$$(23) \quad \alpha_w = 129.62 - (0.5 * 27.333 * 3) = 88.62.$$

Using the yield function parameters, the Stage II primal PMP problem becomes: (see Figure 3).

$$(24) \text{ Max } (2.98*69)x_W + (2.20*65.9) x_O - (88.62 + 0.5*27.333 x_W) x_W - 109.98 x_O$$

$$\text{subject to } x_W + x_O \leq 5$$



A quick empirical check of the calibration to the base values is performed by calculating the VMP per acre of wheat at 300 acres. If it is close to the VMP (VAP) of oats and converging, the model will calibrate without the additional calibration constraints.

The marginal cost per acre of wheat is:

$$MC_{w=3} = 88.62 + 27.333 * 3 = 170.619$$

$$VMP_{w=3} = 2.98 * 69 - 170.619 = 205.62 - 170.619 = 35.001$$

$$VMP_O = 2.20 * 65.9 - 110 = 144.98 - 110 = 34.98$$

The VMP for wheat at 3 acres of \$35.01 is marginally above the VMP for oats (\$34.98). Thus, the unconstrained PMP model will calibrate within the rounding error of this example.

A Lagrangian Approach to PMP - the Wheat / Oats Farm Example

The farm example on page 64 can be further simplified and made consistent with the analytical derivation by only considering the gross margins. GM Wheat = \$76, GM Oats = \$35.

The LP problem is:

$$\begin{aligned} \text{Max} \quad & 76 X_w + 35 X_o \\ \text{subject to} \quad & X_w + X_o \leq 5 \\ & X_w \leq 3 + \varepsilon \\ & X_o \leq 2 + \varepsilon \end{aligned}$$

When we run this problem we will get $X_w = 3 + \varepsilon$, and $X_o = 2 - \varepsilon$. The land constraint will be binding with a shadow value of $\lambda_1 = 35$. The other binding constraint will be the first calibration constraint with a shadow value λ_2 of $(76 - 35) = 41$.

$$\text{Define a PMP cost function } f(X) = 0.5 \gamma X^2. \quad f'(X) = \gamma X$$

$$\begin{aligned} \text{From the PMP theory } \lambda_2 = f'(X) = \gamma X = 41 \text{ when } X = 3. \\ \text{therefore } \gamma = 41/3 = 13.666. \end{aligned}$$

$$\text{The new PMP problem is: } \text{Max} \quad 76 X_w + 35 X_o - 0.5 (13.66) X_w^2$$

$$\text{subject to} \quad X_w + X_o \leq 5$$

The Lagrangian for this problem is:

$$L = 76 X_w + 35 X_o - 0.5 (13.66) X_w^2 + \phi (5 - X_w - X_o)$$

The first order conditions are:

$$\frac{\partial L}{\partial X_w} = 76 - 13.666 X_w - \phi = 0$$

$$\frac{\partial L}{\partial X_o} = 35 - \phi = 0$$

$$\frac{\partial L}{\partial \phi} = 5 - X_w - X_o = 0$$

Solving the second FOC, yields $\phi = 35$. Substituting this into the first FOC results in $76 - 13.666 X_w - 35 = 0$

$$\text{Therefore} \quad X_w = 41 / 13.666 = 3.0.$$

$$\text{Substituting this into the third FOC yields } 5 - 3 - X_o = 0, \quad X_o = 2.$$

NOTE. The PMP problem calibrates at exactly the required crop production acres.

This numerical example shows that PMP models can be calibrated using simple methods. The three stage process and calculation of the parameters is easily programmable as a single process using GAMS/MINOS. Thus, given the initial data and specifications, the PMP model is automatically calibrated in the time it takes to solve an LP and QP solution for the model.

The basic PMP model specified in (24) calibrates in all aspects. That is, the optimal solution, binding constraints, objective function value and dual values will all be within rounding error of the original LP in (16) that is constrained by the calibration constraints. Despite all the notation, the basic concept of PMP is numerically simple and easy to solve automatically on desktop computers.

An alternative calibration approach assumes that the cost per acre is constant, but the yield per acre declines with increased acres to any given crop. It can be shown that this assumption has the equivalent effect on the profit function as increasing costs. Both specifications can be justified from an agronomic point of view. The cost calibration approach is more common than the yield method. The original PMP article (Howitt 1995) uses a yield based calibration method due to intransigent reviewers. The Gams code for the Yolo model without the artificial labor and water constraints, but calibrated using PMP is found in chapter 11.

Calibration Using Elasticity Estimates

Since the PMP procedure solves for the marginal cost function it also solves for the range of supply elasticities based on the marginal costs. However, as can be seen from equation (18), the marginal cost parameter γ_K depends on the empirical parameters of λ_2 and $\tilde{x}_K$. The resulting supply elasticity is not bounded and thus can assume values for a one period calibration that may be inconsistent with estimates based on a larger representative sample of crop response. The elasticity of supply is the essential measure of how the calibrated PMP model responds to policy changes. Accordingly, and consistent with the philosophy of using the best information available, the modeler must check the equilibrium elasticities implied by the calibrated cost functions, and if reliable parameters are available, use the prior information on elasticities to calibrate the model. The PMP marginal cost slope is calibrated against prior econometric estimates, but also reflects the conditions that are present in the base year model conditions. Modelers should be aware that using an elasticity based on prior econometric estimates to calculate the adjustment value does not ensure a positive net return. Net returns over variable costs should be checked after the adjustment factor is calculated.

The supply elasticity based on prior econometric estimates is defined as:

$$\eta_s = \frac{\delta q}{\delta p} \frac{P}{Q}$$

Using the assumption of a constant per acre yield and the usual marginal cost supply function specification, the elasticity can be rewritten in terms of crop land allocations as:

$$(25) \quad \eta_s = \frac{\delta x}{\delta c} \frac{P}{x^*} \quad \text{or} \quad \frac{\eta_s x^*}{P} = \frac{\delta x}{\delta c}$$

Since the nonlinear cost is $0.5 \gamma x^2$, the supply function (marginal cost function) is γx the change

in marginal cost with output is $\frac{\delta c}{\delta x} = \gamma$ therefore $\frac{\delta x}{\delta c} = \frac{1}{\gamma}$

$$\text{and } \frac{\eta_s x^*}{P} = \frac{1}{\gamma}$$

Therefore (26)

$$\gamma = \frac{P}{\eta_s x^*}$$

To check the supply elasticity implied by the unrestricted PMP calibration (26) is reformulated as:

$$(27) \quad \eta_s = \frac{P}{x^* \gamma}$$

If the implied elasticities are outside the expected range that is normally from 0.2 to 2.0, they should be recalculated using equation (25) and the prior elasticities. The cost function intercept parameter α is solved from the average cost equation as in equation (19) except that the elasticity based γ is used.

Calibrating Marginal Crops

A valid objection to the simple PMP specification in (24) is that we assume an increasing cost of production/acre for the more profitable unconstrained crops x_k , but the marginal crops x_m that are constrained by resources are assumed to have constant production costs per acre.

Calibrating the marginal crops (x_m) with increasing cost functions requires additional empirical information. The independent variables, as x_k are termed, use both the constrained resource opportunity cost (λ_1) and their own calibration dual (λ_2) (Figure 2) to solve for the yield function parameters implied by the observed crop allocations. However, the marginal crops (x_m) have no binding calibration constraint, and thus cannot empirically differentiate marginal and average cost at the observed calibration acreage, using the minimal LP data set specified.

Clearly some additional data on the marginal cost function for this group of crops is needed. For cost function calibration, the best additional data comes from prior elasticities of substitution. Since we are now changing the opportunity cost of the restricting resources by changing the costs of the marginal crops, we will have to adjust all the PMP λ_2 values. We use a prior elasticity of supply to calculate the adjustment factor.

Defining the adjustment factor Adj as depending on the slope of the PMP cost function:

$$(28) \quad Adj = \frac{1}{2} \gamma x^* = \frac{P}{2\eta_s}$$

And defining the slope from equation (26) and the prior elasticity yields the second term in (28).

Now we redefine the PMP values for the non-marginal crops as:

$$\lambda_{2i} = \lambda_{2i} + Adj$$

We can now calculate the PMP cost function values of α and γ using the adjusted values and the average costs from the data set.

Returning to the simple pedagogical example in equation (24) and Figure 3, the stage 1 calibrated problem is run exactly as before. One of the important pieces of information from the optimal solution of the stage 1 problem is which activities are in the x_k and x_m groups. The modeler is unlikely to know this beforehand.

In the example, let us assume that the *a priori* information on the elasticity of supply for oats is that it is 2.25. Using equation (28) for the adjustment term (adj_m), the adjustment term for Oats is:

$$(29) \quad Adj_o = \lambda_{2o} = 2.20 * 65.9 / 2 * 2.25 = 28.996$$

Note that this adjustment factor is per acre, so instead of the price per unit product used in the normal elasticity formula, we have to use total revenue (price* yield) per acre, hence (2.20*65.9) which is the price of oats times the yield per acre for oats.

This Adj value now plays the role of λ_2 for the marginal crops, in this example Oats. The residual dual value on land set by the oat cropping activity is reduced accordingly by 28.996 from 35 and becomes 6.004.

Note, in practice it is easy to use low elasticities for the marginal crops, although their nature would lead one to assume a highly elastic supply. From equation (28) it can be seen that there is no bound on the value of Adj, and that small elasticities can lead to large Adj values that in turn lead to negative shadow values and resulting problems in calibration.

The PMP dual on Wheat (λ_{2w}) must also be increased by this same amount to ensure the first order conditions hold. The new value for

$$(30) \quad \lambda_{2w} = 41.0 + 28.996 = 69.996$$

The calculations for the cost coefficients in (22) and (23) are now applied to all activities, both marginal (x_m) and independent (x_k). Note that the adjusted λ_2 values are used for the independent activities and the Adj value based on the prior data is used for the marginal crops.

The PMP problem given the information on marginal yields for the oat crop is defined using the new λ_2 values for both Wheat and Oats in equations (22) and (23) :

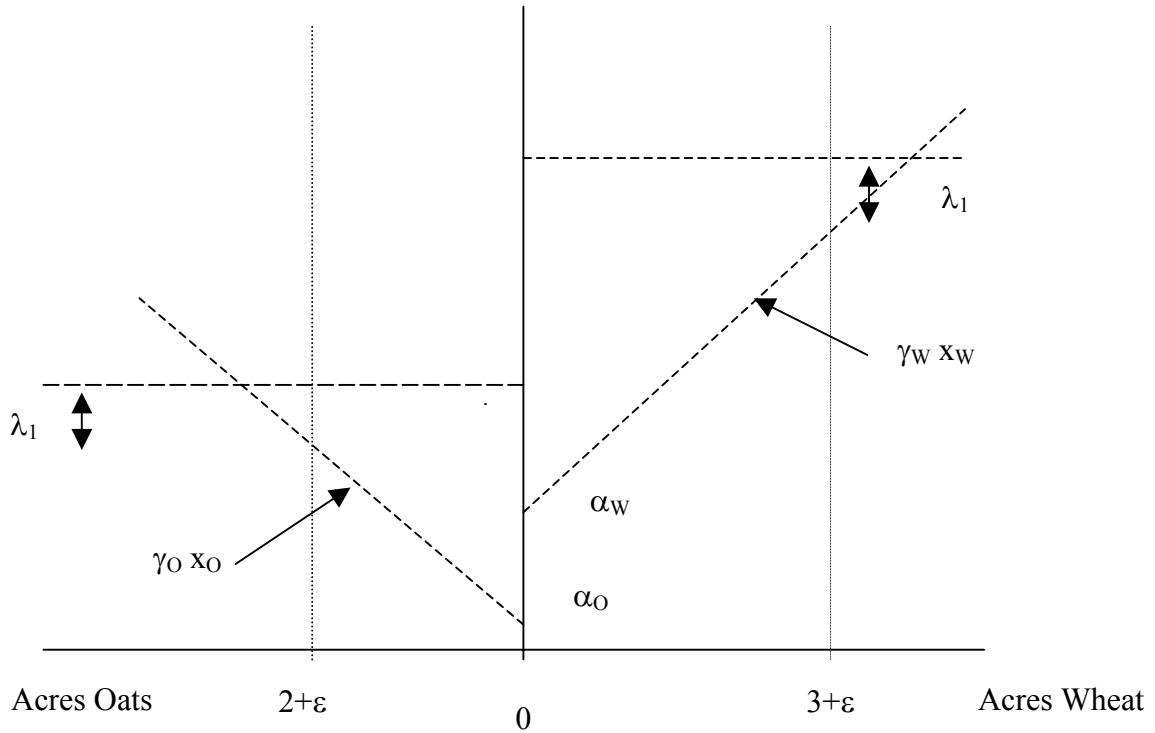
$$(31) \quad \text{Max } (2.98 * 69 - (59.624 + 0.5 * 46.664 * x_w)) * x_w$$

$$+ (2.20*65.9 - (80.984 + 0.5*28.996*x_O)) * x_O$$

$$\text{subject to } x_W + x_O \leq 5$$

The problem is shown in Figure 4.

Figure 4.



The calibration acreage can be checked by calculating the VMP for each crop at the calibration acreages of $\tilde{x}_W = 3$ and $\tilde{x}_O = 2$.

$$(32) \text{ (i) } VMP_W \Big|_{\tilde{x}_W=3} = 2.98*69 - (59.624 + 46.664*3) = 6.004$$

$$\text{(ii) } VMP_O \Big|_{\tilde{x}_O=2} = 2.20*65.9 - (80.984 + 28.996*2) = 6.004$$

Since the VMPs are equal to each other and also equal to the new opportunity cost of land, the PMP model with the new cost functions will calibrate arbitrarily close to the base year acreages.

The resulting model will calibrate acreage allocation and input use, and the objective function value precisely. However, the dual value on land will be lower reflecting the additional, and presumably more accurate, data on the marginal cost of the marginal crops obtained through the elasticities of supply.

Policy Modeling with PMP

The purpose of most programming models is to analyze the impact of quantitative policy scenarios which take the form of changes in prices, technology, or constraints on the system. The policy response of the model can be characterized by its response to sensitivity analysis and changes in constraints.

Advantages of the PMP specification are not only the automatic calibrating feature, but also its ability to respond smoothly to policy scenarios. Paris (1993) shows that the input demand functions and output supply functions obtained by parameterizing a PMP problem satisfy the Hicksian conditions for the competitive firm. In addition, the input demand and supply functions are continuous and differentiable with respect to prices, costs, and right hand side quantities. At the point of a change in basis the supply and demand functions are not differentiable. This is in contrast to LP or stepwise problems, where the dual values, and sometimes the optimal solution are unchanged by parameterization until there is a discrete change in basis, when they jump discontinuously to a new level. This unsatisfactory situation is illustrated by the parameterization of the borrowing limit in the Ohio farm problem in problem set III.

The ability to represent policies by constraint structures is important. The PMP formulation has the property that the nonlinear calibration can take place at any level of aggregation. That is, one can nest an LP sub-component within the quadratic objective function and obtain the optimum solution to the full problem. An example of this is used in technology selection where a specification that causes discrete choices may be appropriate. Suppose a given regional commodity can be produced by a combination of five alternative linear technologies, whose aggregate output has a common supply function. The PMP can calibrate the supply function while a nested LP problem selects the optimal set of linear technology levels that make up the aggregate supply (Hatchett *et al.*, 1991).

Since the intersection of the convex sets of constraints for the main problem and the convex nested sub-problem is itself convex, then the optimal solution to the nested LP sub-problem will be unchanged when the main problem is calibrated by replacing the calibration constraints with quadratic PMP cost functions. The calibrating functions can thus be introduced at any level of the linear model. In some cases, the available data on base year values will dictate the calibration level. Ideally, the level of calibration would be determined by the properties of the production functions, as in the example of linear irrigation technology selection. The PMP approach does not replace all linear cost functions with equivalent quadratic specifications, but only replaces those that data or theory suggest are best modeled as nonlinear.

If the modeler has prior information on the nature of yield externalities and rotational effects between crops, they can be explicitly incorporated by specifying cross crop yield interaction coefficients in equations (13) and (14). The PMP yield slope coefficient matrix is positive definite, $k \neq k$, and of rank k . Without the cross crop effects the matrix is diagonal.

Resource using activities such as fodder crops consumed on the farm may be specified with zero valued objective function coefficients. Where an activity is not resource using, but merely acts as a transfer between other activities, there is no empirical basis or need to modify the objective function coefficients.

References

- Bauer, S. and H. Kasnakoglu. "Non Linear Programming Models for Sector Policy Analysis." *Economic Modelling*, July 1990:275-290.
- Day, R. H. "Recursive Programming and the Production of Supply." *Agricultural Supply Functions*, Heady *et al.*, Iowa State University Press, 1961.
- Hatchett, S. A., G. L. Horner, and R. E. Howitt. "A Regional Mathematical Programming Model to Assess Drainage Control Policies." Chapter 24, pp. 465-489. In *The Economics and Management of Water and Drainage in Agriculture*, Eds., A. Dinar and D. Zilberman. Kluwer, Boston, 1991.
- Hazell P.B.R. and R. D. Norton. *Mathematical Programming for Economic Analysis in Agriculture*, MacMillan Co., New York, 1986.
- Heckelei T, H.P. Witzke, and W. Henrichsmeyer. "Agricultural Sector Modelling and Policy Information Systems" Wissenschaftsverlag Vauk- Kiel KG. 2001
- Howitt, R. E. "Positive Mathematical Programming" *American Journal of Agricultural Economics* 77 (May 1995) :329-342.
- McCarl, B. A. "Cropping Activities in Agricultural Sector Models: A Methodological Proposal." *American Journal of Agricultural Economics* 64:768-771, 1982.
- Meister, A. D., C. C. Chen, and E. O. Heady. *Quadratic Programming Models Applied to Agricultural Policies*, Iowa State University Press, 1978.
- Paris. Q. "PQP, PMP, Parametric Programming, and Comparative Statics." Notes for Ag Econ 253. Department of Agricultural Economics, University of California, Davis, California, 1995.
- Peach, T. *Interpreting Ricardo*. Cambridge University Press, Cambridge. 1993.
- US Bureau of Reclamation, September 1997. Central Valley Project Improvement Act, Draft Programmatic Environmental Impact Statement. Volume 8.

VI Using Nonlinear Models for Policy Analysis

Duality and Parameterization in Nonlinear Models

Deriving the Dual for a Q. P. Problem

Given a vector of outputs, the price dependent demands are defined as:

$$(1) \quad P = \phi + Dx \quad \begin{array}{l} \phi = n \times 1 \text{ vector of intercepts} \\ D = n \times n \text{ negative definite matrix.} \end{array}$$

Monopoly Problem

$$(2) \quad \begin{array}{ll} \text{Max} & \phi'x + x'Dx - \omega'x \\ \text{s.t.} & Ax \leq b \end{array}$$

Deriving Duals for Nonlinear Problems

Given a nonlinear primal problem the dual problem can be derived using the following steps.

1. Set up the problem as a primal Lagrangian. (Here a maximization.)
2. Apply the first Kuhn Tucker condition, $\partial L / \partial x \leq 0$, which yields the constraints for the dual problem.
3. Apply the second Kuhn Tucker condition $\left(\frac{\partial L}{\partial x} \right) x = 0$. Rearrange the equation to obtain $\lambda'Ax$ on the left hand side. Substitute the expression for $\lambda'Ax$ back into the primal objective function. Multiply out and simplify to obtain the dual objective function.

The logic is that by taking the primal first order conditions and objective function, and by substitution, expressing them in terms of prices and costs, we obtain the equivalent dual problem.

Applying this procedure to problem (1) we get the following expressions.

Form a Lagrangian for the problem

$$(3) \quad L = \phi'x + x'Dx - \omega'x + \lambda'(b - Ax)$$

$$= \phi'x + x'Dx - \omega'x + \lambda'b - \lambda'Ax$$

Apply Kuhn Tucker (KT) Conditions $\frac{\partial L}{\partial x} \leq 0 \quad \left(\frac{\partial L}{\partial x} \right) x = 0$

$$(4) \quad \frac{\partial L}{\partial x} = \phi' + 2x'D - \omega' - \lambda'A \leq 0 \quad \text{Transposing}$$

$$\phi + 2Dx - \omega - A'\lambda \leq 0 \quad \text{or}$$

$$(5) \quad A'\lambda \geq \phi + 2Dx - \omega$$

Note equation (5) are necessary conditions in terms of prices, variable costs and imputed costs and thus (5) becomes the dual problem constraint set.

Now we derive the dual objective function.

Using the second KT condition $\frac{\partial L}{\partial x} x = 0$ yields the following condition.

$$(6) \quad (\phi' + 2x'D - \omega' - \lambda'A) x = 0$$

$$\therefore \phi'x + 2x'Dx - \omega'x - \lambda'Ax = 0$$

at the optimum the constraint holds exactly, therefore:

$$(7) \quad \phi'x + 2x'Dx - \omega'x = \lambda'Ax \quad \text{Now plug this into the Lagrangian (3) to get:}$$

$$(8) \quad L = \phi'x + x'Dx - \omega'x - (\phi'x + 2x'Dx - \omega'x) + \lambda'b$$

cancelling

$$(9) \quad L = \phi'x + x'Dx - \omega'x - \phi'x - 2x'Dx + \omega'x + \lambda'b \\ = -x'Dx + \lambda'b$$

The Dual Problem to the Primal Monopoly problem (2) is:

$$(10) \quad \begin{array}{ll} \text{Min} & \lambda'b - x'Dx \\ \text{s.t.} & A'\lambda \geq \phi + 2Dx - \omega \end{array} \quad \lambda \geq 0$$

Economic Interpretation

Note: the Dual objective function has both primal and dual variables in it.

$$(11) \quad \lambda'b \quad \text{— same interpretation as in LP}$$

— the opportunity cost of firm's resources b.

What about the $-x'Dx$ term ?

The Monopoly rent for $x = (\text{Price}(x) - \text{Marginal Revenue}(x))x$ therefore

$$(12) \text{ Monopoly rent} = (\phi' + x'D - \phi' - 2x'D)x \\ = -x'Dx$$

Therefore **the dual objective function in (10) minimizes the sum of imputed resource value and monopoly rent:**

Dual Constraints

$$(13) \quad A'\lambda + \omega \geq \phi + 2x'D$$

(14) $A'\lambda = nx1$ vector of opportunity costs of production of the vector of outputs x .

Since total revenue is $\phi'x + x'Dx$ and total cost is $\omega'x$,

$$(15) \quad A'\lambda + \omega = \text{Marginal opportunity cost} + \text{Marginal cash cost}$$

$$\phi + 2x'D = nx1 \text{ vector of marginal revenue.}$$

Therefore the Dual Monopoly Constraint says:

“The sum of marginal costs of production must be equal to or greater than the marginal revenue for all outputs x .”

Resource Dual Values Under QP

For the binding constraints, substitute the basis matrix B (invertible) into equation (13), for A (non-invertible since not square). This gives a formula for λ 's on binding constraints; on slack constraints $\lambda=0$.

$$(16) \quad B'\lambda = \phi + 2x'D - \omega$$

$$\lambda = B'^{-1}\phi + 2B'^{-1}Dx - B'^{-1}\omega$$

Note: In QP, λ is a continuous function of x , unlike LP where the dual vector is $\lambda = c'_B B^{-1}$

Parameterizing Nonlinear Problems

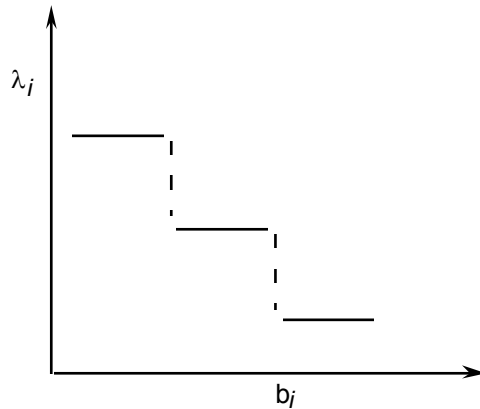
Since the dual values for the primal Linear programming problem are:

$$(17) \quad \lambda = c'_B B^{-1}.$$

This implies that a particular dual value λ_i of b_i is constant until the basis B^{-1} changes.

Note: c_B is a vector of linear net revenues

LP - Parameterization of b_i



The quadratic primal monopolists' problem, however, has a dual of:

$$(18) \quad \lambda = B'^{-1} \phi + 2 B'^{-1} D x - B'^{-1} \omega \quad \text{or}$$

$$(19) \quad \lambda = B'^{-1} (\phi - \omega) + 2 B'^{-1} D x$$

where α and D are the intercepts and slopes of the demands for x_i , and the vector c is defined here as the constant marginal cost/unit of producing x .

Point

The dual λ_i is now a linear function of x , therefore as b_i changes and x changes, the dual will change. However, the "intercept" $B'^{-1} (\phi - \omega)$ will change when the basis changes as will the "slope" $2 B'^{-1} D$ of the dual function.

Condensing the notation by defining $\mu \equiv B'^{-1}(\phi - \omega)$ and $\xi \equiv 2B'^{-1}D$, the dual for the monopoly problem becomes

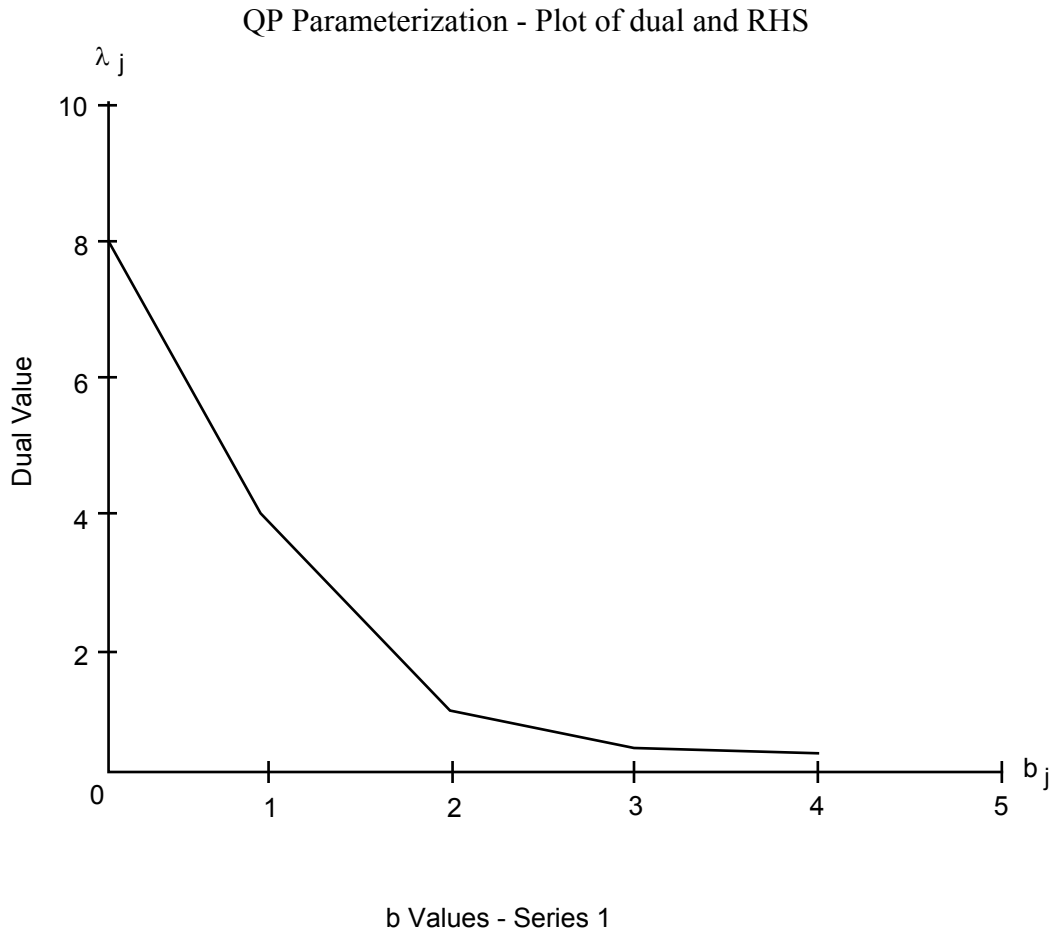
$$(20) \quad \lambda = \mu + \xi x$$

Example

The following simple Quadratic problem illustrates the nonlinear dual.

$$\begin{aligned} \text{Min} \quad & \phi'x + 1/2 x'Dx \\ \text{Subject to} \quad & Ax \leq b \quad x \geq 0 \end{aligned}$$

where $\phi' = [-8 \quad -6 \quad -4]$ and $D = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 0 \\ 2 & 0 & 2 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}$
and b is parameterized over the values $0 \rightarrow 4$.



Comment

As one would expect from the equations (19) and (20) on the previous page, the quadratic dual is a continuous linear function of b_j within a basis, since $x = B^{-1}b$. With a change of basis, the slope of the linear function changes discretely. Equation (19) shows that both the intercept and slope of the dual value function changes with a change in the basis B .

Incorporating Endogenous Supply and Demand Prices

Linear Programs assume constant prices and costs. In reality:

- a) Marginal costs are rarely constant.
- b) Output prices are only constant for individual firms. Analysis which is performed on a regional, national, or commodity basis should have prices that change with changes in the solution.

CASE: I Changes in Output Price Only

Assuming that we are given, or have estimated the parameters of a linear demand function, we can relate the quantity of output demanded q with its price p .

$q = a + Sp$ where: p and q are vectors of prices and quantities.
the parameter (a) is a vector of positive slope intercepts.
 S is a negative definite matrix of demand slopes and cross demand effects.

Note: By definition, a matrix S is negative definite if the scalar product $k'Sk < 0$ for all non-zero values in a conformable vector k .

Since we are modeling the aggregate outcome of individual farmer behavior, we are interested in the opposite effect – i.e., how output levels q affect the price received. This assumption implies that farmers are so small in their individual output that they are price takers. In addition, for most agricultural crops, the farmer has to commit to purchasing the inputs before it is clear what the price will be at harvest. Therefore we invert the demand function to get it in a price dependent form.

$$p = -S^{-1}a + S^{-1}q \quad \text{or} \\ p = \phi + Dq \quad \text{where } \phi \equiv -S^{-1}a \quad \text{and} \quad D \equiv S^{-1}$$

Assuming a constant yield per acre for the moment, we can replace the output quantity (q) by the number of acres allocated to a crop (x) and substitute the resulting expression for price into the objective function. The price endogenous objective function differs with the assumptions on the objectives of the decision maker. There are two main specifications.

- a) Monopolist (Price Manipulator)
- b) Perfect Competition (Price Taker)

(a). Monopoly

Assume a monopolist faces a set of linear price dependent demand functions for the vector of outputs x .

$$(1) \text{ Demand System } \quad P = f + Dx \quad (D \text{ is a symmetric negative definitive matrix})$$

If the monopolist has a constant marginal cost of production vector c the net revenue objective function will be:

$$(2) \text{ Max } J = p'x - w'x \quad \text{substituting in (1)}$$

$$(3) J = (f + Dx)'x - w'x \quad \text{or:}$$

$$(4) \quad \begin{array}{ll} \text{Max } J = f'x + x'Dx - w'x \\ \text{subject to} \quad Ax \leq b \quad x \geq 0 \end{array}$$

Unconstrained Equilibrium

Take the derivative $\frac{\delta J}{\delta x}$ set it = 0 and transpose it. In this problem there are no

binding resources, the constraints are all slack, therefore:

Marginal Revenue = Marginal Cost

$$(5) \quad \begin{array}{ccc} f + 2Dx & = & w \\ \neq & & \neq \end{array}$$

$$\begin{array}{ccc} \text{Vector of Marginal} & = & \text{Vector of} \\ \text{Revenues} & & \text{Marginal Costs} \end{array}$$

Monopoly rent is defined as the difference between total revenue and total cost.

$$\begin{aligned} \text{Monopoly rent} &= p'x - w'x \\ &= (f + Dx)'x - (f + 2Dx)'x \\ &\quad \neq \quad \quad \quad \neq \\ &\quad p \quad \quad \quad \text{MR since monopolist produces where MR=MC} \end{aligned}$$

$$\text{Monopoly Rent} = -x'Dx$$

Note: The monopoly rent $-x'Dx$ is a positive scalar value since D is a negative definite matrix.

Constrained Monopoly Equilibrium

The monopolist is now constrained by a vector of fixed inputs b and the linear technology matrix A . The Lagrangian now becomes

$$(6) \quad \text{Max } J = f(x) + xDx - w(x) + \lambda(b - Ax)$$

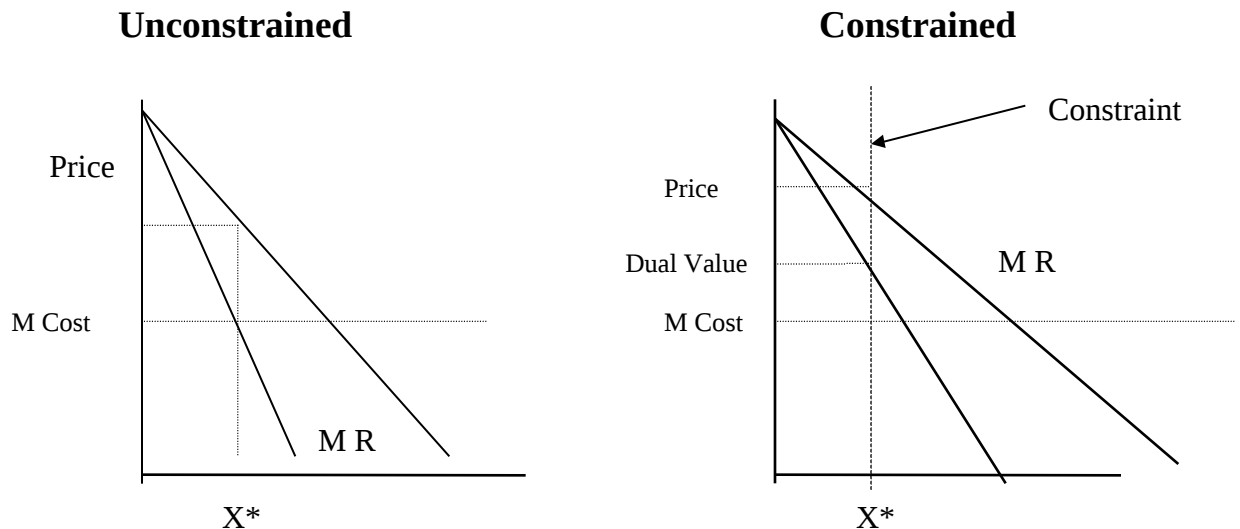
The first order optimum conditions are:

$$(7) \quad \frac{\partial J}{\partial x} = 0 \quad (\text{Set equal to zero})$$

$$(8) \quad f' + 2Dx - w' = A\lambda$$

That is, the difference between the marginal revenue and marginal cost of an output is the sum of shadow values of the inputs used to produce it.

Graph for the Single Product x_i



(b) Aggregated Perfectly Competitive Case

Perfect competition is defined by the following unconstrained equilibrium condition.

$$(9) \quad \frac{\partial z}{\partial x} = 0 \quad (\text{set} = 0 \text{ and transpose}), \text{ yields} \quad \begin{matrix} f' & + & Dx & = & w \\ \neq & & & = & \neq \\ \text{Price} & & & & \text{Marginal Cost} \end{matrix}$$

The perfectly competitive objective function is obtained by integrating the optimal marginal condition from (9), namely.

$$\int (\phi + Dx - \omega) dx = \phi'x + 1/2 x'Dx - \omega'x$$

Accordingly, we specify a different objective function that satisfies the marginal conditions for **unconstrained** perfect competition.

$$(10) \text{ Max } z = (\phi + 1/2 Dx)'x - \omega'x$$

A good question is: Why is the one half in the objective function multiplying the slope parameter ? The answer is that a Perfectly Competitive market is defined by its marginal conditions, so to correctly define the objective function, we have to start with the marginal conditions and derive the objective function. Essentially we have to ask, what objective function would an optimizing decision maker have had to result in these first order conditions ? We therefore start with the marginal conditions and integrate them to obtain the objective function.

The Constrained Perfect Competition Problem.

$$(11) \text{ Max } z = \phi'x + 1/2 x'Dx - \omega'x$$

Subject to $Ax \leq b \quad x \geq 0$

The perfectly competitive objective function also maximizes the sum of consumer's surplus and producer's quasi-rent (producer surplus)

$$\begin{aligned} (12) \quad z &= (\phi + 1/2 Dx)'x - \omega'x && \text{(add and subtract } 1/2 x'Dx) \\ &= (\phi + Dx)'x - \omega'x - 1/2 x'Dx && \text{(substitute in for price } p) \\ &= p'x - \omega'x - 1/2 x'Dx \\ &= (P-\omega)'x - 1/2 x'Dx \end{aligned}$$

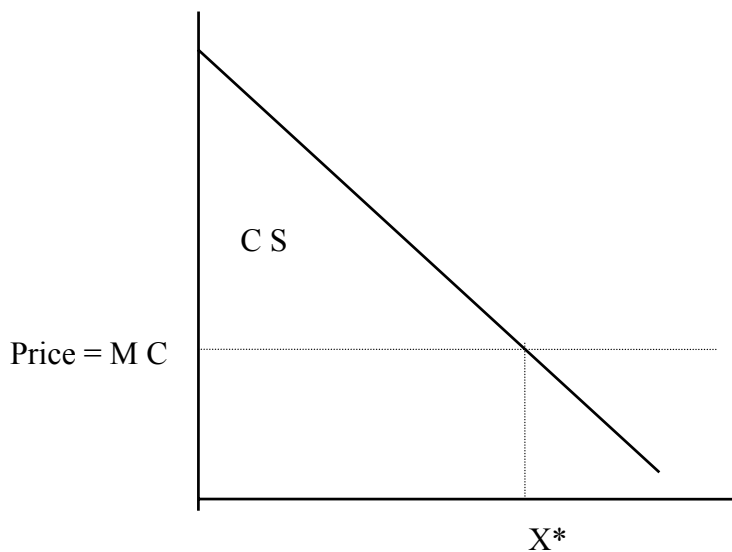
Since Price minus Variable Cost is defined as “Quasi Rent” or producer's surplus, the left hand side term $(P-\omega)'x$ is equal to quasi rent. Since the marginal cost is defined as a constant value “ ω ”, the quasi rent only occurs because of the constraint on the amount of product sold.

What about the right hand side term $- 1/2 x'Dx$?

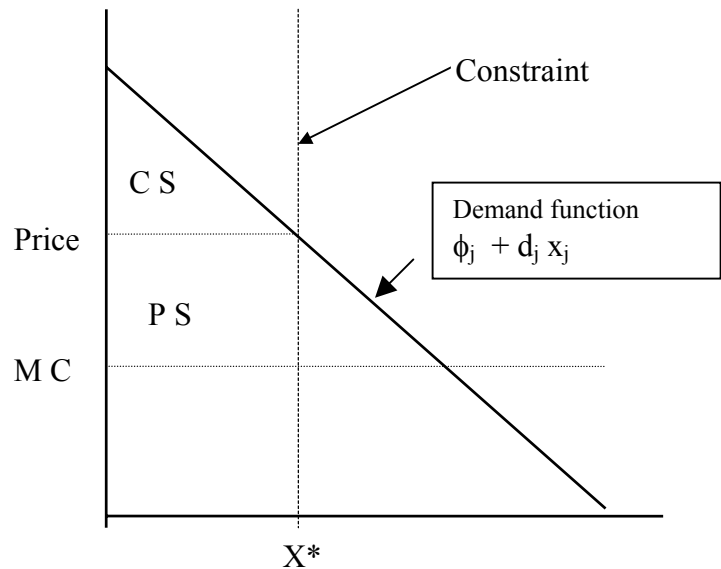
$$\begin{aligned} (13) \quad -1/2 x'Dx &= 1/2 x'(-Dx) && \text{(add and subtract } \phi) \\ &= 1/2 x'(\phi - \phi - Dx) && \text{(substitute in for price } p^*) \\ &= 1/2 x'(\phi - P^*) \\ &= \text{consumer's surplus} && \text{(see figure)} \end{aligned}$$

Thus equation (9) (a) Satisfies competitive marginal conditions
(b) Maximizes aggregate net social benefits.

Unconstrained perfect Competition



Constrained Perfect Competition



Summary.

Unconstrained Perfect Competition

Producer Surplus (p.s.) = 0

Consumer Surplus (c.s.) = $1/2(\phi_i - p_i)x_i$

Constrained Perfect Competition

Producer Surplus (p.s.) =
quasi-rent attributed the fixed input b_j
 $(p_i - \omega_i) x_i$

Consumer Surplus (c.s.) = $1/2(\phi_i - p_i)x_i$

CASE II Aggregate Perfect Competition—Endogenous Prices and Costs

In this specification the marginal costs is no longer constant and there is a linear supply function (Marginal Cost) as well as endogenous demand prices.

Given the price dependent demand function and a linear marginal cost (supply) function below

$$\text{Price} = \phi + Dx$$

D is negative definite

$$(15) \quad \text{Marginal Cost} = \alpha + \Gamma x$$

Γ is positive definite.

The unconstrained problem is therefore.

$$(16) \quad \text{Max } J = \phi'x + 1/2x'Dx - \alpha'x - 1/2 x'\Gamma x$$

$$(17) \quad \frac{\partial J}{\partial x} = \phi' + x'D - \alpha' - x'\Gamma \quad \underset{=}{\text{set } 0}$$

As expected for perfect competition—price = marginal cost.

The Interpretation of the Objective Function

Trick #1. Add and subtract $1/2x'Dx$ and $1/2x'\Gamma x$ to (16) to yield:

$$(18) \quad J = \phi'x + x'Dx - \alpha'x - x'\Gamma x - 1/2x'Dx + 1/2x'\Gamma x$$

$$(19) \quad J = (\phi + Dx)'x - (\alpha + \Gamma x)'x - 1/2 x'Dx + 1/2 x'\Gamma x$$

But from (17) we see that the first two terms cancel out (Price = Marginal cost) at the optimum therefore at the optimum the objective function is:

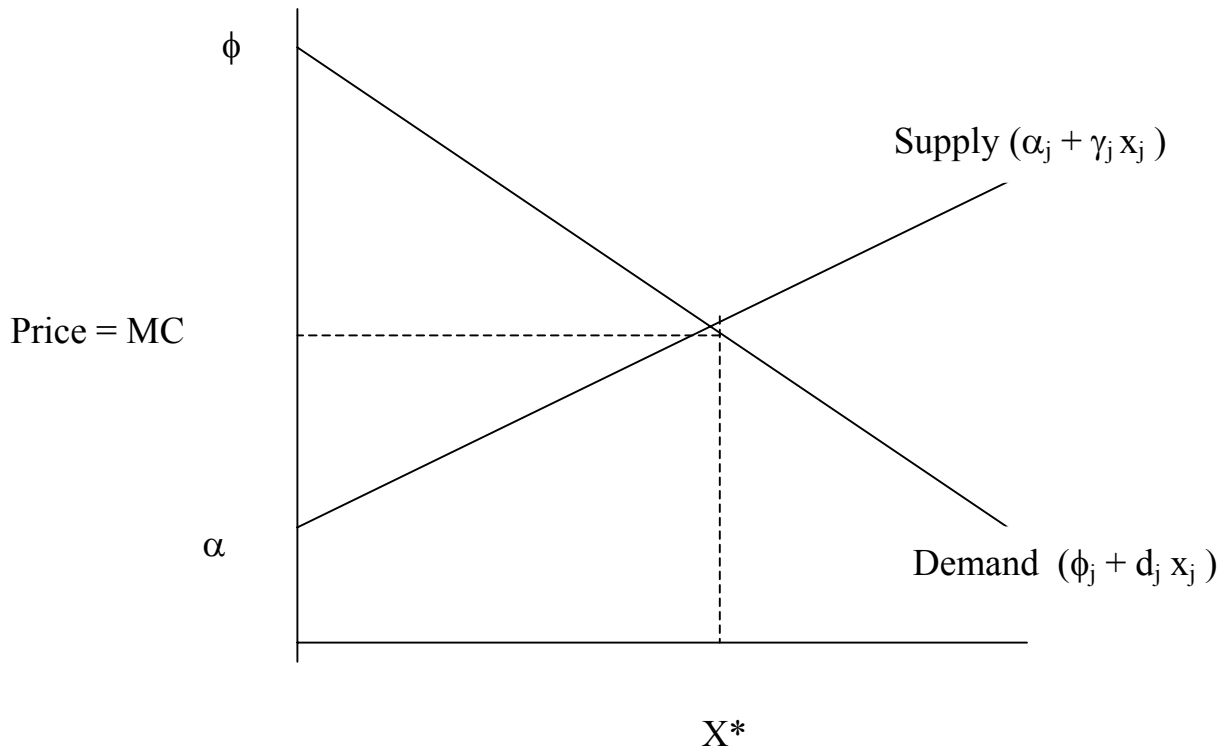
$$(20) \quad J = -1/2x'Dx + 1/2x'\Gamma x$$

From (13) we see that the first term $-1/2 x'Dx$ is equal to consumer's surplus. The second term is changed by trick #2—factoring out x and adding and subtracting α to yield:

$$(21) \quad \begin{aligned} 1/2 x'\Gamma x &= 1/2 (\alpha + \Gamma x - \alpha)' x \\ &= 1/2 (MC - \alpha) x \end{aligned}$$

Since price equals marginal cost, $1/2 x'\Gamma x$ is one half the area above the marginal cost intercept and below the price line as shown below.

Perfect Competition–Endogenous Price and Cost



CASE III Inter-regional Trade Models

Empirical trade models can be solved using this same objective function expanded to several regions. The effect of transport costs and tariffs can be added to the supply functions to solve for changes in trade policies or conditions. The seminal paper in this area is by Takyama and Judge (1964).

There are several different ways of setting up the inter-regional trade problem, but the simplest method to show is to extend the quantity dependent supply / demand concept above to J regions that are linked by trade which results in a trading cost of ω_{ij} per unit commodity traded from region “i” to region “j”. The resulting problem solves a wide range of trade problems

Quantity Dependent Optimal Inter-regional Trade Specification

$$\begin{aligned}
 \text{Max } F(.) &= \sum_{j=1}^J (\phi_j + 0.5d_j) x d_j - \sum_{i=1}^I (\alpha_i + 0.5\gamma_i) x s_i - \sum_i \sum_j c_{ij} x_{ij} \\
 \text{s.t. } x s_i &= \sum_j x_{ij} \\
 x d_j &= \sum_i x_{ij} \\
 x_{ij} &\geq 0
 \end{aligned}$$

The objective function maximizes the definite integral under the demand and supply function for each region in terms of the post trade quantities demanded and supplied in each region. The costs of trading between regions is deducted to yield the net social benefit of trade shown in the two region diagram. The adding up constraints ensure that the quantities demanded and supplied in each region balance.

The specification of the regional trade problem is an excellent illustration of the efficiency and beauty of dual specifications. The primal specification above solves optimally, but it is a bit more complicated than needed. Since the decision variable is the quantity of product traded between regions, the cost of trading is explicitly defined in the objective function, and the aggregate quantities are generated by the summing up constraints.

An alternative to using quantity dependent supplies and demands, is to formulate the dual of the quantity dependent problem that can be termed the price dependent form. The price dependent specification solves the problem with two simple equations that use the standard quantity dependent demand and supply functions, and instead of solving for $i*j$ quantities traded, the price based model solves for the $i + j$ set of equilibrium prices.

The two equations are the producer and consumer surplus objective function, and the first order price condition for trade. Returning to the original quantity dependent demand on page 76 we have:

$$x d_j = a_j + s_j p d_j$$

We also define a similar quantity dependent supply function

$$x s_i = b_i + g_i p s_i$$

The Price Dependent Interregional Trade Problem

$$\begin{aligned} \text{Max } F(.) &= \sum_{j=1}^J (a_j + 0.5 s_j p d_j) p d_j - \sum_{i=1}^I (b_i + 0.5 g_i p s_i) p s_i \\ \text{s.t.} \quad & p d_j - p s_i - c_{ij} \leq 0 \end{aligned}$$

The quantities traded are generated as the dual values to the (i * j) interregional pricing constraints. This is an example of the complementary slackness principle working for us. From the complementary slackness principle we know that if the trade price constraint is slack, that is the supply and transport cost exceed the demand price, then the quantity traded will be zero. The corollary is that the dual value when the trade price constraint is binding is the quantity traded. Note that all the information from the optimal solution of the primal problem is also obtained from solving the simpler dual problem.

An empirical example of the Gams code for these problems is found in chapter 10. By running both problems it can be seen that the results are identical.

Calibrating Demands Using Elasticity Estimates

Often the modeler is faced with the need to derive parameters for demands and supplies when the only data available is the equilibrium price and quantity in the base year for the model, and an estimate of the elasticity from an previous econometric study. There is a very simple derivation that enables one to calibrate the slope and intercept coefficients using the elasticity. Assume that a demand can be specified as linear in form used above.

$$pd = \phi + d x$$

Recall that that demand elasticity is defined as:

$$\eta = \frac{\delta q}{\delta p} \frac{p}{q}$$

If you have a base value for $q = 2.9$, $p = 173$, and the elasticity = -0.6 the values can be substituted in to the above formula to solve for the slope of the demand function

$$-0.6 = \frac{\delta q}{\delta p} \frac{173}{2.9}$$

The slope of the price dependent linear demand function that results is:

$$\frac{\delta p}{\delta q} = -99.42$$

Substituting this value back into the original price dependent demand equation results in an intercept value of 461.3

The resulting calibrated demand curve that will yield a price of \$173 at a quantity of 2.9 and a point elasticity of -0.6 has the form:

$$pd = 461.3 - 99.42 q$$

The concept of calibrating models against prior econometric elasticity estimates is well illustrated in the Central Valley Production Model (CVPM) that has been developed by S Hatchett in the consulting firm CH2M - Hill. The CVP model is currently used to analyze the economic impacts of large reallocations of water between agricultural, environmental and urban interests in California. The model use both demand, supply and substitution elasticities to reflect the economic impacts of changes in water availability and cost in California.

VII Incorporating Risk and Uncertainty

In our problem specifications we have implicitly assumed that the parameters in the problem are known and constant. For example, the objective function $c'x$ assumes a deterministic vector c of prices or costs. We can make this more realistic in two ways:

- Assume the elements of c are not known with certainty, but they are stochastic with known distributions.

$$c \sim N(\bar{c}, \Sigma_c)$$

where Σ_c is an $n \times n$ variance/covariance matrix of revenues.

- The decision maker's objective function values both the expected return and its variance.

Question

If the vector of net revenues c is distributed $c \sim N(\bar{c}, \Sigma_c)$, what is the distribution of the objective function $c'x$, where x is a non-stochastic vector.

Answer

Since x is a deterministic linear operator

- Expected Value $E(c'x) = \bar{c}'x$
- Variance of $c'x = E\{c'x - E(c'x)\}^2$

$$\begin{aligned} \text{Var}(c'x) &= E\{[c'x - E(c'x)][c'x - E(c'x)]\} \text{ (Inner Product)} \\ &= E\{x'[c - E(c)][c - E(c)]'x\} \text{ (Inner Product of an Outer Product)} \\ &= x'E\{[c - E(c)][c - E(c)]'\}x \text{ (Taking Expectation of stochastic terms)} \\ &= x'\Sigma_c x \text{ (by covariance matrix definition)} \end{aligned}$$

Point

If $c \sim N(\bar{c}, \Sigma_c)$ then

$$c'x \sim N(\bar{c}'x, x'\Sigma_c x) \quad \text{where } x \text{ is deterministic.}$$

If the decision maker is risk neutral, the objective function has $\bar{c}$ substituted for c . But more likely, the decision maker is risk averse.

If the degree of aversion to income variance can be measured by a parameter " ρ " then the problem becomes:

$$\begin{array}{ccc} \max z = \bar{c}'x & - & \rho x' \Sigma_c x \\ \uparrow & & \uparrow \\ \text{(expected revenue vector)} & & \text{(variance of revenue)} \end{array}$$

Subject to $Ax \leq b \quad x \geq 0$.

The Effect of Uncertainty and Risk Aversion

This nonlinear risk cost will have several different effects on the optimal solution of the problem.

- (i) The optimal solution will show more diversification to offset risk.
- (ii) Since the problem is no longer linear in x , some of the x_j 's may have interior optima and will not be restricted by binding constraints. In this case, there will be more x_j activities than constraints.
- (iii) Note that ρ is a scalar, since the variance of a vector product ($c'x$) is a scalar. " ρ " measures the "cost of risk" to the decision maker. The variance of the return from the portfolio $x' \Sigma_c x$ and the expected return $\bar{c}'x$ are both changed by changing the proportions of x_i in the portfolio.

The point is demonstrated by focusing on a single x_i , and denoting the sum vector of derivatives that result from the covariance by the short hand expression $\frac{\delta(\text{var})}{\delta x_i}$.

$$\begin{aligned} \frac{\delta z}{\delta x_i} &= \frac{\delta(\bar{c}'x)}{\delta x_i} - \rho \frac{\delta(\text{var})}{\delta x_i} = 0 \quad (\text{Set equal to zero for an unconstrained optimum}) \\ \therefore \rho \frac{\delta(\text{var})}{\delta x_i} &= \frac{\delta(\bar{c}'x)}{\delta x_i} \\ \therefore \rho &= \frac{\delta(\bar{c}'x)}{\delta x_i} \frac{\delta x_i}{\delta(\text{var})} = \frac{\delta(\bar{c}'x)}{\delta(\text{var})} \end{aligned}$$

That is ρ is equal to the marginal rate of tradeoff between expected income and variance at the optimum.

Measuring Risk Aversion- E/V Analysis

In the past section, note that the risk aversion parameter $\rho = \frac{1}{\lambda}$ where λ = opportunity cost of constraint when you minimize the variance of the objective function. The risk aversion parameter can be calculated for an individual by solving the following E/V problem:

$$\begin{aligned} \text{Min } v &= x' \Sigma_c x \\ \text{Subject to } \bar{c}'x &\geq e^* \end{aligned}$$

Where: $x' \Sigma_c x$ = Variance of revenue, $\bar{c}'x$ = Expected revenue
 e^* = The value chosen for the minimum expected revenue

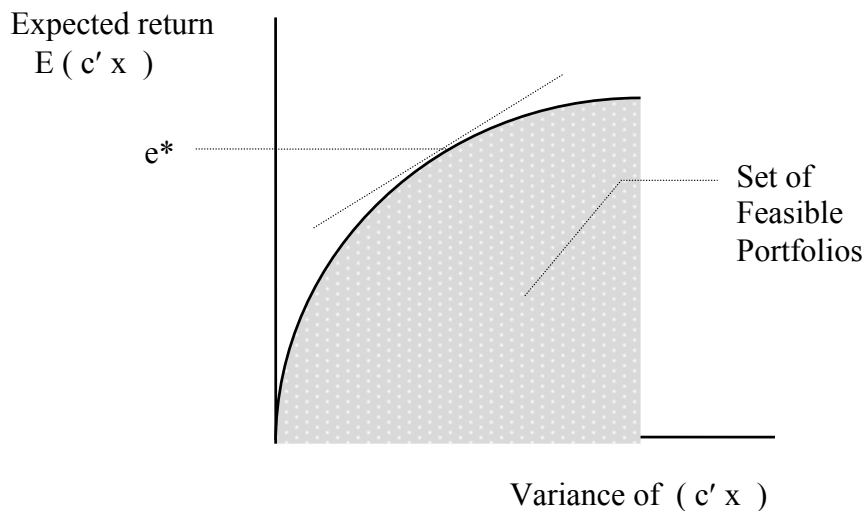
Expressing this problem as a Lagrangian

$$L = \text{Var} + \lambda (e^* - \bar{c}'x)$$

$$\frac{\partial L}{\partial x_i} = \frac{\partial(\text{var})}{\partial x_i} - \lambda \frac{\partial(\bar{c}'x)}{\partial x_i} \text{ set } = 0 \quad \therefore \lambda = \frac{\partial(\text{Var})}{\partial(\bar{c}'x)} = 1/\rho$$

The model can generate "E.V." frontiers which show those output combinations that give the best combination of expected revenue (E) and variance of revenue (V). (See graph below)

E . V. Frontier



The EV frontier is generated by the following steps

- (1) $\min \text{var} (c'x) \quad \text{s.t.} \quad \bar{c}'x \geq e^* \quad \text{for a range of } e^* \text{ values}$
- (2) plot results

A simple example of the Gams code for the E/V problem is found in chapter 10.

Note: There is a linear approximation to the mean/variance objective function called MOTAD. (see Hazell and Norton pp. 86-90). With the growth in nonlinear algorithms and computing power, this linear approximation is rarely needed.

Uncertainty in the Constraints: Chance Constrained Programming

The constraints of a problem sometimes are not known with certainty. Often the quantity available of input resources (b_i 's) is uncertain, for farming this may be reflected by the distribution of growing season length, rainfall or seasonal labor availability. The a_{ij} technological coefficients may also be stochastic, but this more complex case is set aside for the moment.

Case Considered

- a) Some or all of the right hand side b_i 's are stochastic. Their distribution is known.
- b) The problem decision maker has specified that the problem solution must have the uncertain constraint satisfied for a known proportion (α) of the time. That is the probability that the constraint is satisfied is specified.

Probability Review Any normal random variable can be transformed to a variable that has a "Standard Normal Distribution" -N (0,1). whose cumulative probabilities are calculated and tabulated (z distribution)

$$\text{if } b_i \sim N(\bar{b}_i, \sigma_{b_i}^2) \text{ then } \frac{b_i - \bar{b}_i}{\sigma_{b_i}} \sim N(0,1)$$

Also the probability that a z distributed random variable exceeds a specified value is equal to one minus the cumulative probability at that value.

Case: A Single Right Hand Side Value b_i is Normally Distributed

- (1) Given the problem $\text{Max } c'x$
s.t. $Ax \leq b$

the i^{th} row in the constraints is written:

$$(2) \quad \sum_{j=1}^n a_{ij}x_j \leq b_i$$

But if b_i is stochastic and we want this constraint to hold with probability α , it is rewritten as:

$$(3) \quad \text{Prob} \left\{ b_i \geq \sum_{j=1}^n a_{ij}x_j \right\} \geq \alpha$$

This is the probability that the constraint is slack, or just holds. Note that the right hand side of the expression in the brackets $a_{ij}x_j$ is deterministic but changes with changes in x .

Step 1

If we require the probability that a normally distributed random variable will be equal to or bigger than a number we need to calculate a value where cumulative probability is α , our specified level. It is quick and convenient to convert the distribution to an equivalent standard normal and look up the probabilities in a table.

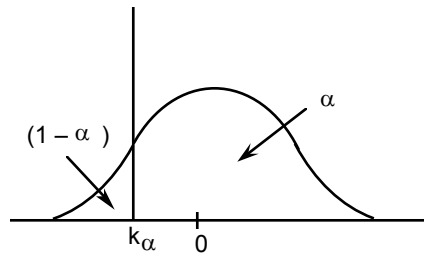
Note: Z tables are usually set up so that:

- 1) They only tabulate half the distribution, so you have to add or subtract .5 from the cumulative probability.
- 2) They give the cumulative probability in the "tail" of the distribution.

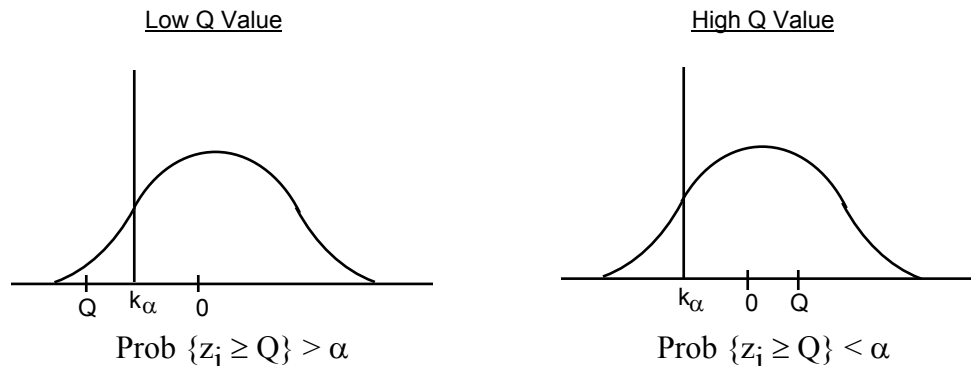
- (4) If the variable z_i is distributed $z_i \sim N(0,1)$

That is, z_i is distributed with a standard normal distribution, the probability that $\{z_i \geq Q\}$ is greater than α holds only if $Q \leq k_\alpha$.

Where the cumulative probability $(1 - \alpha)$ is the area under the curve (density function) in the tail beyond k_α .



Example



Step 2.

- a) Convert the normally distributed b_i to an equivalent standard normal distribution, as above.

$$(5) \quad b_i \text{ is "standardized" to } z_i = \frac{b_i - \bar{b}_i}{\sigma_{bi}}$$

Since we want to put this "standardized" value into the constraint (3) we have to

perform the same b_i standardization transformation on the other $(\sum_{j=1}^n a_{ij}x_j)$ side

of the constraint. Applying the transformation to both sides, constraint (3) is rewritten for the standard normal distribution as equation (6). Note that the left hand term in the brackets is a random variable, while the right hand term is deterministic.

$$(6) \quad \text{Prob} \left\{ \frac{b_i - \bar{b}_i}{\sigma_{bi}} \geq \frac{\sum_{j=1}^n a_{ij}x_j - \bar{b}_i}{\sigma_{bi}} \right\} \geq \alpha$$

Note: a) $\frac{b_i - \bar{b}_i}{\sigma_{bi}} \sim N(0,1)$ That is, it has a standard normal distribution.

b) $\frac{\sum_{j=1}^n a_{ij}x_j - \bar{b}_i}{\sigma_{bi}}$ is composed of fixed and known parameter values and thus is

equivalent to the value Q used above.

From Step 1 and equation (4) we see that $\text{Prob} \{z_i \geq Q_i\} \geq \alpha$,

holds only if $Q_i \leq k_\alpha$. Therefore, expression (6) only holds if:

$$(7) \quad \frac{\sum_{j=1}^n a_{ij}x_j - \bar{b}_i}{\sigma_{bi}} \leq k_\alpha$$

Step 3

Multiplying out (7), we see that (6) only holds if:

$$(8) \quad \sum_{j=1}^n a_{ij}x_j - \bar{b}_i \leq k_{\alpha}\sigma_{b_i} \quad \text{or rewriting (8)}$$

$$(9) \quad \sum_{j=1}^n a_{ij}x_j \leq \bar{b}_i + k_{\alpha}\sigma_{b_i}$$

An intuitive explanation is to think of $k_{\alpha}\sigma_{b_i}$ as a risk adjustment factor that changes the probability of the constraint binding.

For example, if $k_{\alpha}\sigma_{b_i} = 0$, the constraint would bind 50 percent of the time.

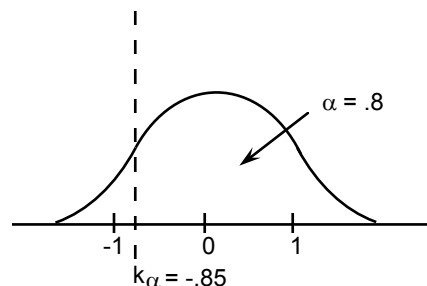
- Note:
- a) The left hand side is the familiar Ax constraint.
 - b) The right hand side is the original b_i value adjusted by a term that is linear and can be calculated from the Z tables and the distribution of b_i .
 - c) The adjusted constraint holds with a probability α .

A Numerical Example

Suppose $b_i \sim N(42,4)$, does a resource requirement $\sum a_{ij}x_j = 40.25$ hold with probability .8? In this example we have a fixed decision variable and want to know if it holds with a given probability

Step 1. Find k_{α} , using the Z tables. Since $\alpha = .8$, greater than .5, we look up $(1 - \alpha)$ which here = .2. The tables give us a value of .85, which we deduce from symmetry is $-.85$. Note the .8 and .85 are **completely** different parameters.

$$\therefore k_{.8} = -.85$$



Step 2. Find "Q" using values from the distribution of b_i

$$"Q" = \frac{\sum_{j=1}^n a_{ij}x_j - \bar{b}_i}{\sigma_{b_i}} = \frac{40.25 - 42}{2} = -.875$$

Step 3. Comparing k_α and Q, we see that $Q < k_\alpha$, i.e., $-.875 < -.85 \therefore$ we know that the constraint value of 40.25 will be satisfied slightly more than 80% of the time.

Uncertainty in the Technical “ a_{ij} ” Coefficients

Case a) Individual a_{ij} 's have a known mean and variance
b) Two or more a_{ij} 's have a non-zero covariance.

Result We use the same probability concept as with the b_i , but the derivation results in nonlinear (quadratic) constraints in x .
For details see Hazell and Norton pp.107-110.

For a more general model specification that combines the mean / variance objective function and chance constraints on the input supply, see Paris (1979)

Paris Q. “Revenue and Cost Uncertainty, generalized Mean- Variance, and Linear Complementarity Problem”. A J A E. 61, 2: (May 1979): 268- 275.

VIII An Introduction to Maximum Entropy

Measuring Information

The PMP cost function in chapter 5 has its quadratic coefficients solved analytically by solving two equations in two unknowns. However, this analytical solution requires that the quadratic cost matrix is specified as strictly diagonal. There is a significant practical problem with the diagonal specification in that it assumes that there are no "cross effects" between the amount of land allocated to crops, apart from the effect on the total land constraint. In economic parlance this assumption requires that there are no substitution or complementarity cost effects between crops grown in the same district or farm. Clearly the almost universal existence of rotations in crop production implies that farmers are well aware of the interdependencies among crops, and use them to stabilize or increase profits. Clearly, the assumption of a diagonal cost matrix is unrealistic. But to calibrate a full matrix of coefficients requires solving an "ill posed" problem in which we are trying to estimate more parameter values than we have data points, an estimation with negative degrees of freedom. As we saw in chapter 2, we cannot solve ill posed problems by inverting matrices. Fortunately there is an alternative method that can obtain parameter estimates for ill posed problems using information theory and the principle of maximum entropy.

Claude Shannon was a giant in information theory. In 1948, he published a paper that proposed a mathematical way of measuring information, and started the information revolution. Shannon noted that information is inextricably linked with the probability of an event about which the signal tells us. Clearly a signal that tells us that an extremely unlikely event, such as a bad earthquake in Davis, has gone from a very low probability of happening to a very high probability has a very high information content. Note, we are not even considering whether the event has any particular value associated with it. Likewise a signal that tells us that a very likely (high probability) event will happen has a low information content.

Shannon proposed several axioms that a measure of information must have, and showed that the only measure of information content that satisfied the axioms is if the information content of a signal is:

- (1) $-\ln(p)$ where p is the prior probability of the event happening.

Shannon extended the definition from single probabilities to discrete distributions and defined the expected information content of a prior distribution $\sum_i p_i$ as the entropy of a distribution:

- (2) $H = -\sum_i p_i \ln p_i$.

It follows that a distribution that has a uniform distribution, in which each event is equally likely, has the highest entropy and the lowest information content. Conversely, a distribution that puts a weight of one on a single outcome and zero on the rest, has an entropy of zero. Remember that, counterintuitively, a distribution that shows that a given

event will occur with probability one, has the highest information content (lowest entropy).

If we specify a set of discrete values over the feasible range of a parameter value termed the **"support set or space"** then multiplying each support value by its associated probability yields and expected parameter estimate, and there is an entropy value for the distribution associated with the parameter value. However, as one would expect with an ill posed problem, there is an infinite set of probabilities that will yield any given parameter value. We have to use the entropy criteria to choose a unique distribution fomr among the infinite set of feasible distributions.

Jaynes , another information pioneer, showed that the distribution with the highest entropy can occur in the most number of ways. The concept of "multiplicity" is similar to frequency or maximum likelihood in conventional estimators. Essentially the distribution with the maximum entropy is the "best" estimator. Maximizing entropy also has another fine property, in that the entropy function in equation (2) has a unique solution at the maximum.

Thus by defining a support space for our parameter and so9lving for the maximum entropy distribution that has an expected value consistent with the data, we can get a unique solution to the ill posed problem, and estimate more parameters than we have observations!

The “Loaded” Dice Example

Here we se how we intuitively calculate six parameters (probabilities) from a single data observation. Jaynes in his 1963 Brandeis lectures on maximum entropy used the example of a game using a six sided dice with the usual values from 1 – 6. Suppose that you know the average value of a large number of independent rolls of the dice. For a given mean value there are an infinite number of combinations of the six probabilities that could have generated the mean value. The problem is ill posed because we are given one data point, the mean value, and we have to estimate six probabilities. The only structural constraint that we have is that the probabilities have to add to one.

Think about a game in which your opponent produces a dice and suggests that it is rolled twenty times, and if the average score is above 4 you win, and if the average score is below 3 they win. With a “fair dice” in which we assume that the probability of each side coming up is even, this is a fair game. If you notice that the rolls of the dice come up consistently with 1,2,or 3, you will alter your initial assumption that the dice is fair, and assume that your opponent is cheating with a dice loaded to favor the low scores. You have just performed an ill-posed estimation of the probabilities.

Using the principle of multiplicity, the distribution that maximizes the entropy is the most likely to be probabilities underlying the dice. If we define the score values as x_i , the dice problem is specified as:

$$\text{Max} \quad H = - \sum_i p_i \ln p_i$$

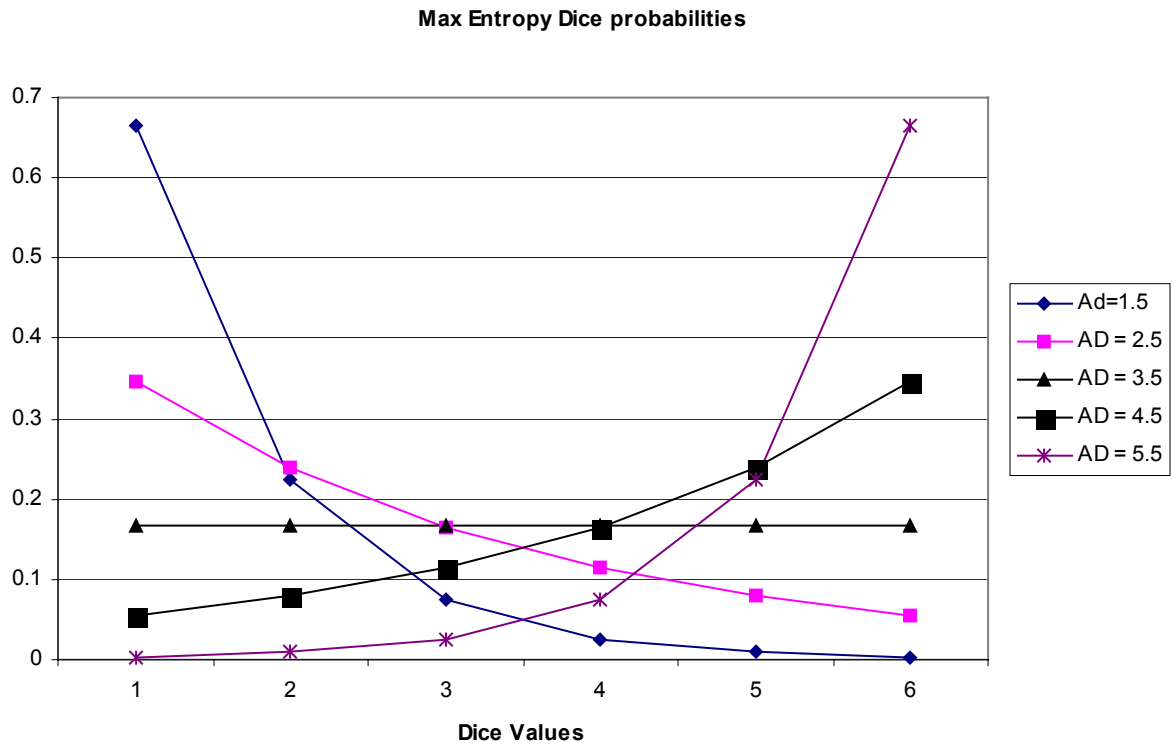
(6)

subject to

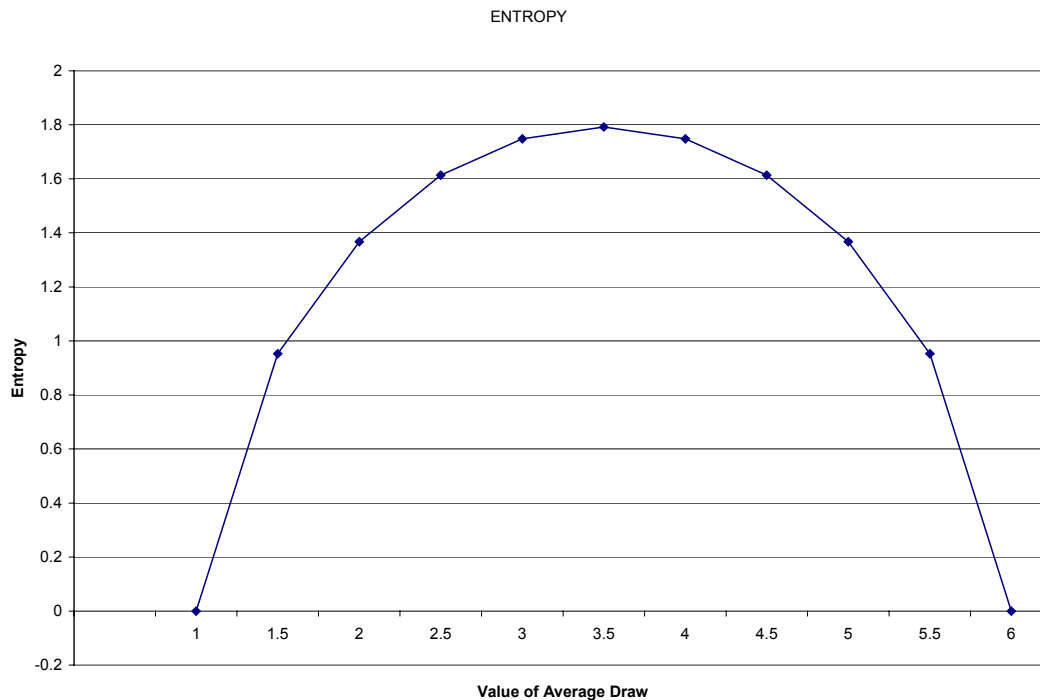
$$\text{Av Score} = \sum_i x_i p_i$$

$$\sum_i p_i = 1, \quad p_i \geq 0$$

The Gams solution to the dice problem is in the website Gams folder. The ME probability results for a range of average scores from 1.5 to 5.5 is plotted below.



The entropy value for different average scores is plotted below, note the unique maximum value.



A Simple Example of Maximum Entropy Parameter Estimation

Assume that we want to estimate two parameters in the simple quadratic cost function:

$$(3) \quad TC = a x + 0.5 b x^2$$

We only have one observation in which we see that the marginal cost is 60 when the output $x = 10$. The data relationship that we have to satisfy is:

$$(4) \quad 60 = a + 10 b.$$

There are an infinite number of parameter values for "a" and "b" that satisfy this relationship.

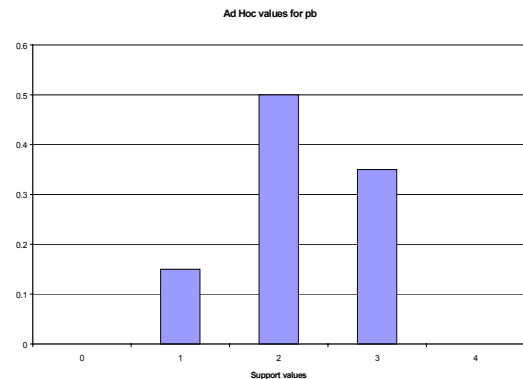
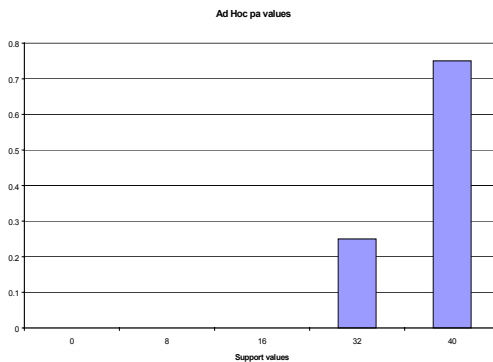
Suppose we consider five discrete values for a support space. If we rule out negative costs, the lower support space is bounded at zero. the upper support space can be defined by the coefficient value that would explain all of the cost when the other coefficient is zero. Using this as a basis for the support values, five evenly distributed support values would be.

$$(5) \quad z_{a_i} = [0, 8, 16, 32, 40] \quad z_{b_i} = [0, 1, 2, 3, 4]$$

A feasible set of probabilities that would solve the equations

$$(6) \quad MC_j = \sum_i z_{a_i} p_{a_i} + (\sum_i z_{b_i} p_{b_i}) x_j$$

$$\sum_i p_{a_i} = 1, \text{ and } \sum_i p_{b_i} = 1, \quad p_{a_i}, p_{b_i} \geq 0$$



is $p_{a_i} = [0, 0, 0, 0.25, 0.75]$ and $p_{b_i} = [0, 0.15, 0.5, 0.35, 0]$

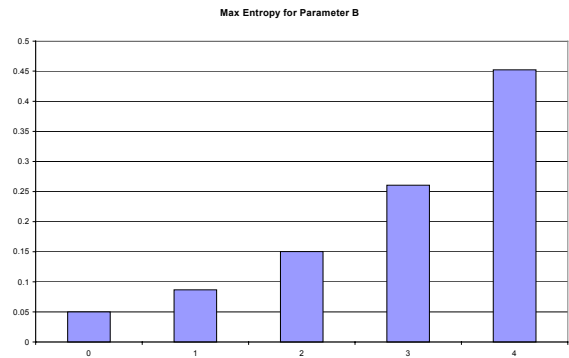
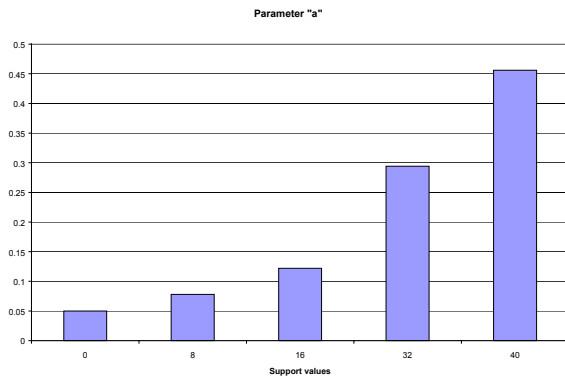
The distributions for this solution are plotted on the histograms above.

The Maximum Entropy Solution

The maximum entropy problem that solves for the two distributions that are most likely to have generated a marginal cost of 60 for an output of 10 is:

$$\begin{aligned}
 (6) \quad \text{Max } H &= - \sum_i p_{a_i} \ln p_{a_i} - \sum_i p_{b_i} \ln p_{b_i} \\
 &\text{subject to} \\
 MC_j &= \sum_i z_{a_i} p_{a_i} + \left(\sum_i z_{b_i} p_{b_i} \right) x_j \\
 \sum_i p_{a_i} &= 1, \text{ and } \sum_i p_{b_i} = 1, \quad p_{a_i}, p_{b_i} \geq 0
 \end{aligned}$$

The maximum entropy (ME) solution to this problem is plotted on the histograms below. Clearly the ME solution satisfies the data constraint without having to use such specialized, and unlikely, distributions as the ad hoc solution above.



The expected parameters that result from these two calculations are shown below

	Ad Hoc Values	Maximum Entropy
E(a)	38.0	30.217
E(b)	2.2	2.978

Defining the Support Space Values (Z values)

Since the z values are “priors” on the parameters that we want to reconstruct they can have a strong influence on the resulting parameters. We need to be aware that the defined z values must be:

1. Feasible for the data constraints that we are going to impose on the support space.
2. Neutral in their effect on the estimate unless we have information that we want to incorporate in the priors.

Feasible Supports.

In the simple parameter and dice example the z values of the dice are clear, and for the cost parameters they were glossed over. When defining a system for generating z values for different models and crop components in the models we need a more formal approach. Feasible z values are generated by defining the z values (say five) as being the product of a single centering parameter value and five z weight parameters.

$$zval(j, p) = zwtg(p) * cval(j).$$

The centering $cval(j)$ is an empirical value the modeler calculates will be feasible for the data set imposed on it. For example, if we are reconstructing a PMP cost function matrix as a function of the acreage planted of a crop, then a feasible value for the marginal cost coefficient is the average cost divided by the base acres.

Non-Informative z values

The $zwtg(p)$ parameters can be thought of as spreading the $cval(j)$ value across a feasible range for estimation. For a diagonal cost parameter that is strictly positive, the range may go from 0, 0.5, 1.0, 1.5, 2.0. For an off-diagonal parameter the $cval(j)$ weight will be reduced, possibly to 0.25 the diagonal weight and centered on 0, with weights of -1.5, -0.75, 0.0, 0.75, 1.5..

This system of generating z values is designed to automatically generate a set of feasible but noninformative prior z values.

Informative Prior z values.

In the Yolo example the z values on the cost slope parameters are defined to be informative, in that their centering value is defined by the model parameters and a prior elasticity value. The elasticity value is entered in the same manner as in the PMP elasticity calibration, and forms a stronger basis for the defined z values. Note this elasticity is based on “scalar” reasoning in the calculation is based on single coefficients and ignores the effect of the off diagonal coefficients and resource constraints, however, as can be seen from the empirical elasticity test, they do give the model an operating range of elasticities.

Adding Curvature Conditions by Cholesky Decompositions

To ensure that the resulting matrix PMP model converges to a stable solution the second order conditions require that the Hessian of the cost function is negative definitive. Since the Hessian is the Zeta matrix, this requires that Zeta is positive definite. Diewert shows that a necessary condition for a matrix to be positive definite is that the diagonal elements of its Cholesky decomposition matrix are positive.

What is a Cholesky decomposition ? Judd(1999) says that we can think of a Cholesky decomposition as being the “square root” of a matrix. The C matrix is a lower triangular matrix L that when post multiplied by its transpose yields the original matrix.

If $Z = L L'$ then L is the Cholesky decomposition.

If Z is a 3 x 3 matrix.

$$Z = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \end{bmatrix}$$

The Cholesky decomposition $L = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$ and

$$Z = L L' = \begin{bmatrix} l_{11}l_{11} & l_{11}l_{21} & l_{11}l_{31} \\ l_{21}l_{11} & l_{21}l_{21} + l_{22}l_{22} & l_{21}l_{31} + l_{22}l_{32} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}l_{31} + l_{32}l_{32} + l_{33}l_{33} \end{bmatrix}$$

This latter expression results in two sets of equations for the diagonal and off-diagonal elements of Zeta.

$$z_{jj} = \sum_{k=1}^j l_{jk}^2 \quad \text{and} \quad z_{ij} = \sum_{k=1}^j l_{ik} l_{jk}$$

by adding these equations as constraints on the entropy problem and constraining the diagonal Cholesky terms to be greater than zero (here > 0.001) we can impose curvature conditions on the Zeta matrix that ensures that the simulated problem will satisfy the second order conditions. The program that implements a ME estimate of the PMP cost matrix for the Yolo model is found in the Gams folder on the website.

Reconstructing Production Function Models

We may have "n" observations over time on production units. For the time being, we will simplify the problem to a single observation of a unit that produces "j" crops each of which has "i" inputs. There is an input subset of restricted, but allocatable inputs such as land or irrigation water. The data set consists of observations on crop price, input price, crop yield, and input use by crop. This data set can be used to define the implicit Leontieff matrices and specify the following calibrated linear programming problem.

$$\begin{aligned}
 (1) \quad & \text{Max } \sum_j p_j y_j x_j - \sum_{ij} \omega_i a_{ij} x_j \\
 \text{s.t.} \quad & Ax \leq b \\
 & Ix \leq \tilde{x} + \varepsilon \\
 & x_j \geq 0
 \end{aligned}$$

The first set of allocatable resource constraints generate the shadow values for those constraints that influence the observed crop and input allocation. The perturbed upper and lower bound calibration constraints ensure that the crop allocation is within ε of the observed data, and in addition, provide measures of the rotational cost interdependence between the crops based on the equi-marginal principle for land allocation.

Before the ME reconstruction program is run, support values have to be defined for each parameter and error term. To ensure that the set of support values spans the feasible solution, we define the support values as the product of a set of five weights and functions of the average Leontieff yield over the data set, and for a particular crop/input combination. The support values for the error terms are defined by positive and negative weights that multiply the left-hand side values of the equation.

Curvature is added by solving for the parameters of the Cholesky decomposition of the quadratic matrix L where $Z = LL'$, and constraining the diagonal Cholesky parameters to be nonnegative, for details see (Paris & Howitt 1999). The quadratic production function for a single crop j as a function of i inputs is defined as:

$$y_j = \alpha' x_j - 0.5 x_j' Z x_j$$

The ME reconstruction problem for a single crop with "i" inputs becomes:

$$\text{Max } \sum_{i,p} p a_{i,p} - \ln p a_{i,p} + \sum_{i,ii,p} p l_{i,ii,p} - \ln p l_{i,ii,p}$$

Subject to :

$$sh_i = \sum_p p a_{i,p} * z a_{i,p} - [\sum_{ii} [\sum_p (p l_{i,ii,p} * z l_{i,ii,p}) * \sum_p (p l_{ii,i,p} * z l_{ii,i,p})] * x_{ii}]$$

$$yld = \sum_p p a_{i,p} * z a_{i,p} - [\sum_i \{ (\sum_{ii} [\sum_p (p l_{i,ii,p} * z l_{i,ii,p}) * \sum_p (p l_{ii,i,p} * z l_{ii,i,p})] * x_{ii}) \} * x_i] / x_{land}$$

$$\sum_p p a_{i,p} = 1, \quad \sum_p p l_{i,ii,p} = 1,$$

The objective function is the usual sum of the entropy measures for the two sets of parameters for the Cholesky decomposition of the quadratic matrix and the vector of linear terms. The first equations are the first order conditions that set the cost share equal to the marginal physical product. If some inputs are restricted and the PMP calibration stage is used, the input cost in the share equation will include the resulting shadow values as well as the nominal input price.

The second set of equations fit the production function to the observations on average yield. While it is not normal in econometric models to include average product equations, the information in this constraint is particularly important for two reasons.

First, information on yields is likely to be the most precisely know by farmers. Second, while the marginal conditions are essential for behavioral analysis, policy models also have to accurately fit the total product to be convincing to policy makers and correctly estimate the total impact on the environment and the regional economy of policy changes. Fitting the model to the integral as well as the marginal conditions improves the policy precision of the model.

The production function parameters are calculated from the entropy results for each crop as:

$$\alpha_i = \sum_p p a_{i,p} * z a_{i,p}$$

$$zeta_{i,ii} = \sum_{i2} [\sum_p (p l_{i,i2,p} * z l_{i,i2,p}) * \sum_p (p l_{ii,i2} * z l_{ii,i2})]$$

The quadratic production problem is defined as:

$$\text{Max } \sum_j p_j f_j(x_{ij}) - \sum_{ij} \omega_j x_{ij}$$

subject to

$$\sum_j x_{ij} \leq b_i$$

$$x_{ij} \geq 0$$

$$\text{where } f_j(x_{ij}) = \alpha'_j x_j - 0.5 x'_j Z x_j$$

Note that the production technology in the objective function and constraints is no longer Leontieff.

Calculating Comparative Static Parameters for the Model

The quadratic production function model has convenient properties for calculating policy parameters. Note that the Hessian of the constrained profit function above is simply:

$$\frac{\partial^2 \Pi}{\partial x_{ij}^2} = Z$$

Calculating the Derived Demands for Inputs

For simplicity, we will use the unconstrained profit function for a single crop.

$$\Pi = p(\alpha'x - 0.5x'Zx) - \omega'x$$

$$\frac{\partial \Pi_j}{\partial x} = p'_j(\alpha - Zx) - \omega = 0$$

$$-Zx = \frac{\omega}{p_j} - \alpha$$

$$x^* = Z^{-1}\alpha - Z^{-1}\frac{\omega}{p_j}$$

$$\text{Define the matrix of demand slopes } G_j = Z^{-1}\frac{1}{p}$$

$$\text{and the vector of intercepts } a_j = Z^{-1}\alpha$$

the system of input demands based on crop j is :

$$x^* = a_j + G_j \omega$$

From the above equations, it is clear that if we can invert the Hessian of the profit function (Z^{-1}) we can calculate the derived demand for each input for each crop as a linear function of input and output price.

Note

- (i) That x^* is an $i \times 1$ vector of the optimal inputs for crop j, and a_j is an $i \times 1$ vector of intercept terms and G_j is an $i \times i$ matrix of derived demand slopes
- (ii) That the demand for a given input i used in crop j is a function of the prices of the other inputs as well as its own price.

The elasticity of demand is based only on the “own price” effect, thus we need to get Gams to only use the i^{th} diagonal elements of DS_j

The elasticity of the input demand follows directly :

$$\text{The elasticity is } \eta_{ij} = G_{j,i,i} \frac{\omega_i}{x_i^*}$$

This elasticity is based on a single crop. For the usual multi-output case we weight the individual crop contribution by their relative resource use to arrive at a weighted elasticity for the resource.

Calculating Supply Functions and Elasticities

Since production is a function of optimal input allocation and we now have the input demands as a function of input and output price, we can derive the output supply function by substituting the optimized input derived demands into the production function and simplify in terms of the output price. Going back to the derived demand and production function formulae:

$$x^* = Z^{-1}\alpha - Z^{-1} \frac{\omega}{p_j}$$

and

$$y_j^* = \alpha' x^* - 0.5 x^{*'} Z x^*$$

$$\text{Defining } r_i = \frac{\omega_i}{p_j}$$

Using the vectors α and r

$$y^* = \alpha' Z^{-1}(r - \alpha) - 0.5 (r - \alpha)' Z^{-1} Z Z^{-1}(r - \alpha)$$

multiplying out and collecting α and r terms yields

$$y^* = 0.5 \alpha' Z^{-1} \alpha - 0.5 r' Z^{-1} r$$

$$y^* = \hat{\alpha} - 0.5 \left(\frac{\omega_1}{p} \dots \frac{\omega_I}{p} \right)' Z^{-1} \left(\frac{\omega_1}{p} \dots \frac{\omega_I}{p} \right)$$

Substituting the expression for y^* into the supply elasticity formula, and separating out “ ω ” and “ p ” we get the supply elasticity expression:

$$\eta_{sj} = \frac{0.5 \alpha' Z^{-1} \alpha}{\sqrt{p} y_j}$$

Calculating Elasticities of Input Substitution

There are many elasticities of substitution with different advantages and disadvantages. To demonstrate that we can obtain crop and input specific elasticities of substitution we use the classic Hicks elasticity of substitution defined by Chambers in (Applied Production Analysis, 1988) on page 31 as:

$$y^* = f(x_1, \dots, x_I)$$

$$\sigma_{1,2} = \frac{-f_1 f_2 (x_1 f_1 + x_2 f_2)}{x_1 x_2 (f_{11} f_2^2 - 2 f_{12} f_1 f_2 + f_{22} f_1^2)}$$

where f_i and f_{ij} are derivatives

Output from the Yolo Model

```

----      324  PARAMETER  DIFF              PERCENT DIFFERENCE IN INPUT
              LAND          WATER              LABOR
R1.ALFA      0.0238947    0.0238946    0.0238946
R1.WHEAT     -0.0099991   -0.0099990   -0.0099989
R1.CORN      -0.0099990   -0.0099990   -0.0099990
R1.TOM       -0.0099990   -0.0099990   -0.0099990

```

```

----      324  PARAMETER  YDIFF              PERCENTAGE DIFFERENCE IN YIELD
              ALFA          WHEAT          CORN          TOM
R1 -2.54829E-8  5.558568E-8 -1.88046E-8 -6.03031E-9

```

```

----      324  PARAMETER  SUPELAS              SUPPLY ELASTICITY
              ALFA          WHEAT          CORN          TOM
R1   1.8559458    0.8040508    1.0985663    0.5755326

```

```

----      324  PARAMETER  DDELAS  CROP SPECIFIC INPUT DEMAND ELASTICITY
              LAND          WATER          LABOR
R1.ALFA      2.5277203    1.2055752    1.3689239
R1.WHEAT      2.1832609    0.6338516    0.3691518
R1.CORN      2.3500945    1.0376875    0.5785391
R1.TOM       1.0477321    0.1712690    1.6733499

```

```

----      324  PARAMETER  DEMELAS              AGGREGATE INPUT DEMAND ELASTICITY
              LAND          WATER          LABOR
R1   2.0916067    0.8465534    1.4426146

```

---- 324 PARAMETER SUB ELASTICITY OF SUBSTITUTION

		WATER	LABOR
ALFA .LAND	1.3291195	1.2590338	
ALFA .WATER		1.2621399	
WHEAT .LAND	0.8346119	0.4626905	
WHEAT .WATER		0.4797180	
CORN .LAND	1.2829549	0.7437865	
CORN .WATER		0.7369215	
TOM .LAND	0.1828788	1.3330871	
TOM .WATER		0.1800782	

---- 324 PARAMETER ZETA PROD MATRIX

		LAND	WATER	LABOR
ALFA .LAND	0.0020656	-0.0000580	-0.0000613	
ALFA .WATER	-0.0000580	0.0001221	0.0000016	
ALFA .LABOR	-0.0000613	0.0000016	0.0000257	
WHEAT .LAND	0.0011258	0.0000408	0.0000573	
WHEAT .WATER	0.0000408	0.0003063	0.0000297	
WHEAT .LABOR	0.0000573	0.0000297	0.0002101	
CORN .LAND	0.0022263	-0.0000174	0.0000227	
CORN .WATER	-0.0000174	0.0002478	0.0000130	
CORN .LABOR	0.0000227	0.0000130	0.0001730	
TOM .LAND	0.0295139	0.0023333	0.0000889	
TOM .WATER	0.0023333	0.0036581	0.0000757	
TOM .LABOR	0.0000889	0.0000757	0.0000135	

---- 324 PARAMETER ALPHA PROD LINEAR

		LAND	WATER	LABOR
R1.ALFA	2.1538230	0.4178334	0.1905927	
R1.WHEAT	1.5242699	0.5242465	0.3938534	
R1.CORN	1.6476534	0.3968946	0.2847649	
R1.TOM	16.8443240	4.6199548	0.4301530	

For further reading see:

Golan, Judge & Miller "Maximum Entropy Econometrics" Wiley 1996

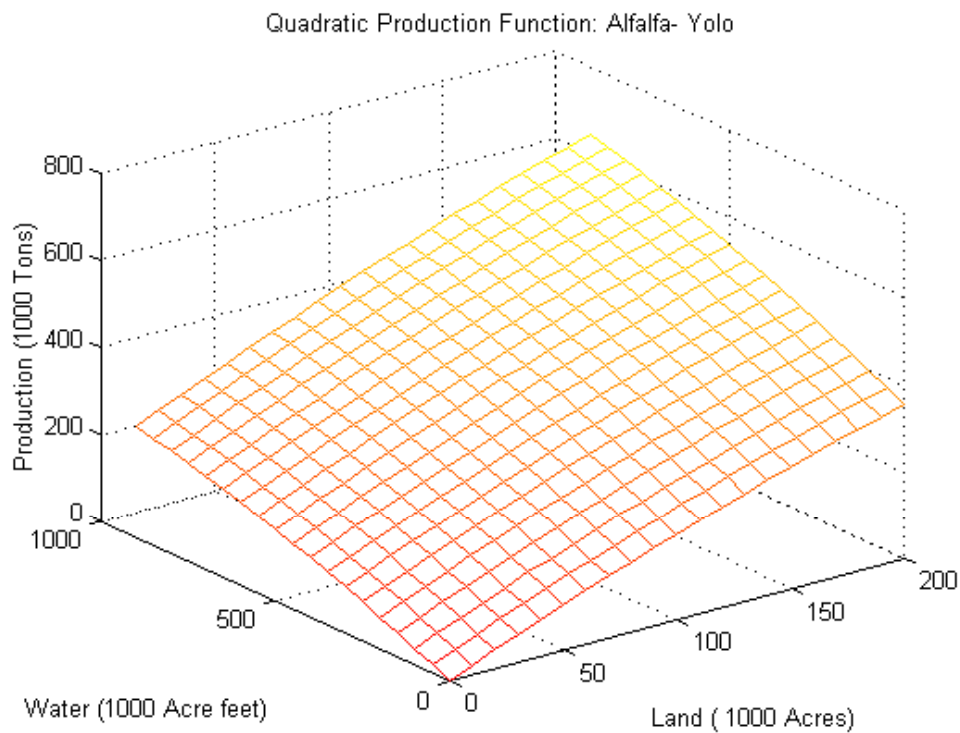
Mittlehammer, Judge & Miller "Econometric Foundations" 2000.

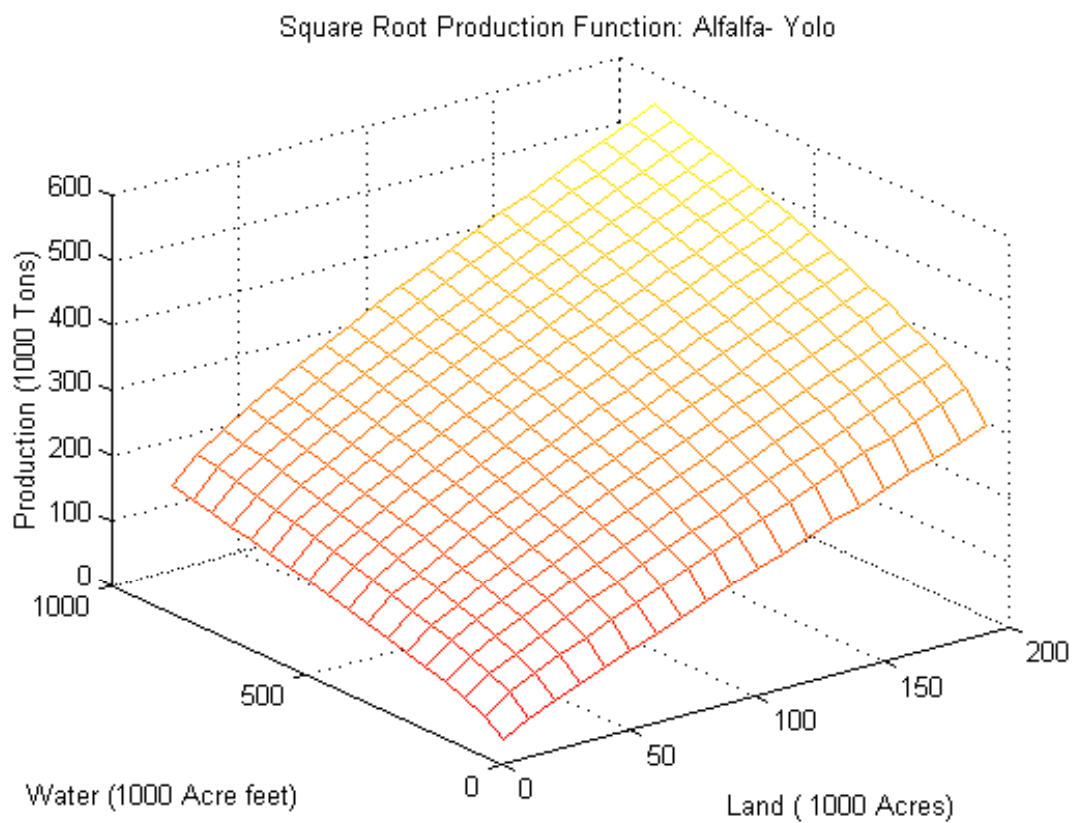
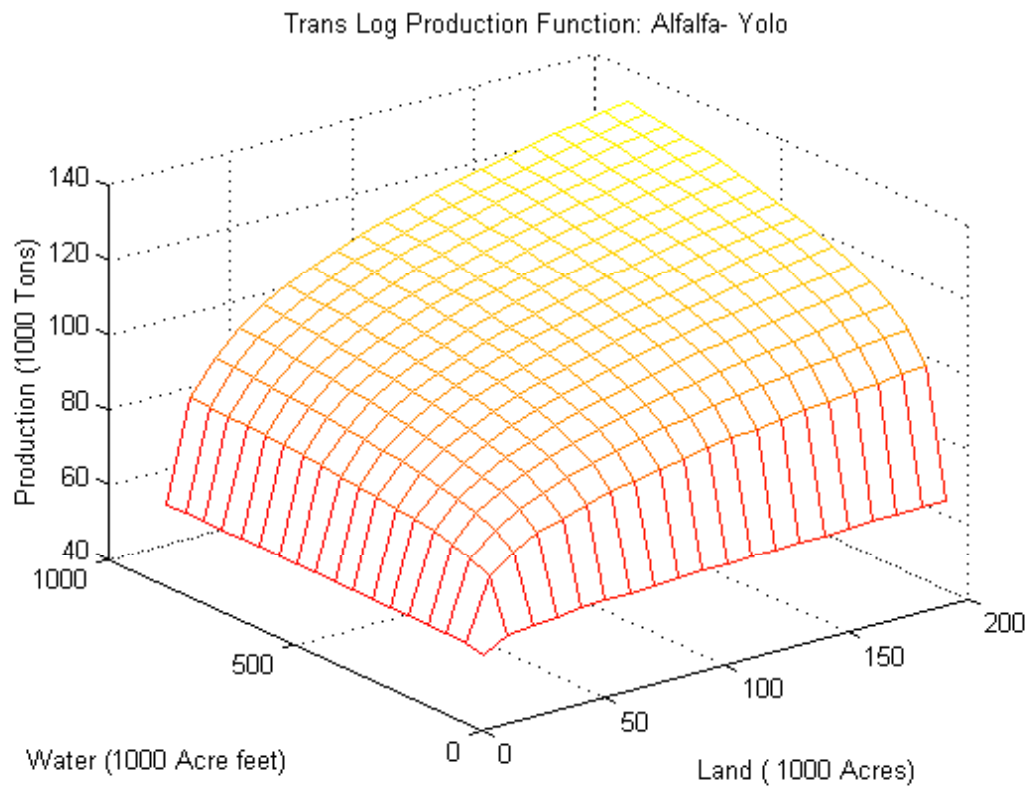
Paris & Howitt AJAE 80: Feb 1998, 124- 138

Using Alternative Functional forms for the Production Function

Throughout this analysis we have used the quadratic production function. Two other functional forms that are widely used are the Generalized Leontieff and Translog production functions. As would be expected, the ME empirical reconstruction methods outlined in this chapter apply equally well to these production functions.

For illustration, I show the production surface for all three production specifications for the Yolo model simplified to the two inputs of land and water.





IX Nonlinear Optimization Methods

Mathematics For Nonlinear Optimization

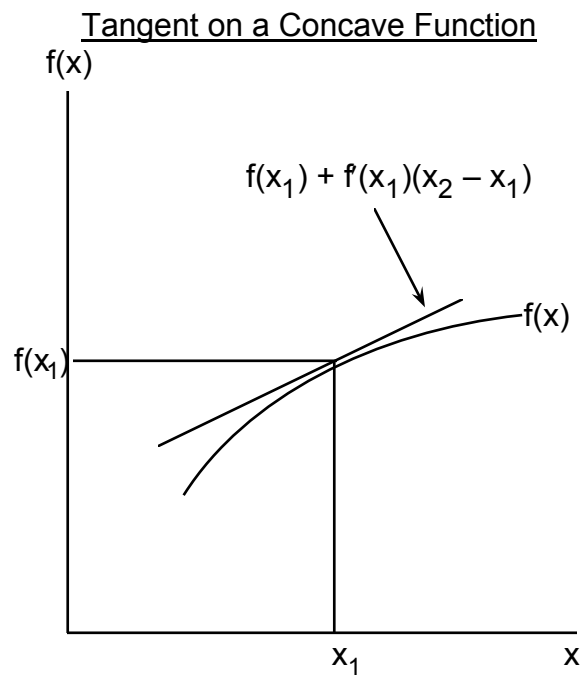
Concave Functions and Convex Sets

Concave Function A function $f(x)$ defined on a convex set Ω is strictly concave if for every $x_1, x_2 \in \Omega$ and every $\lambda, 0 \leq \lambda \leq 1$

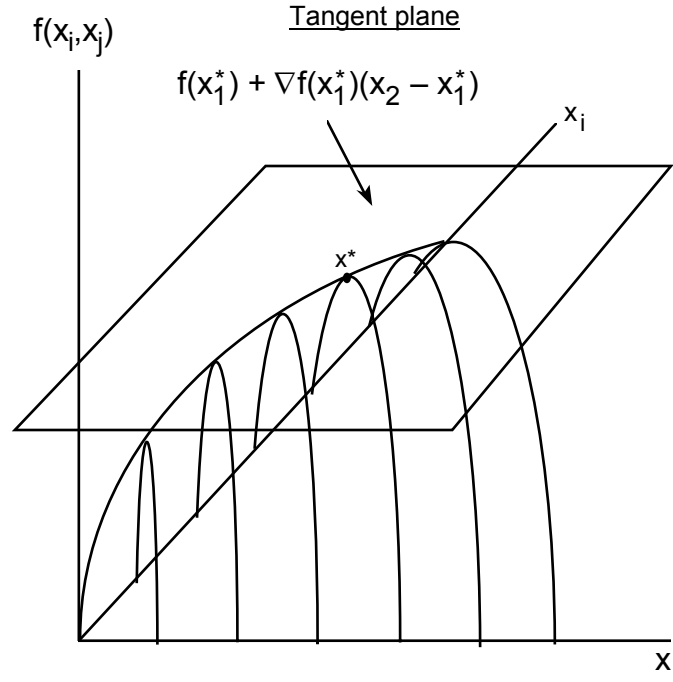
$$f(\lambda x_1 + (1-\lambda) x_2) > \lambda f(x_1) + (1-\lambda) f(x_2)$$

The tangent line of a scalar valued function $f(x)$ at x_1 is

$$f(x_1) + f'(x_1)(x_2 - x_1)$$



It follows that the tangent plane for a vector valued function $f(x)$, at a point in n space, say x_1 can be expressed as $f(x_1) + \nabla f(x_1)(x_2 - x_1)$ for $x_2 \in \mathbb{R}^n$.



Note: If $f(x)$ is a concave function on the convex set Ω , then the set $\Omega = \{x: x \in \Omega, f(x) \geq c\}$ is convex for every real number c . This is a method for defining a nonlinear constraint set.

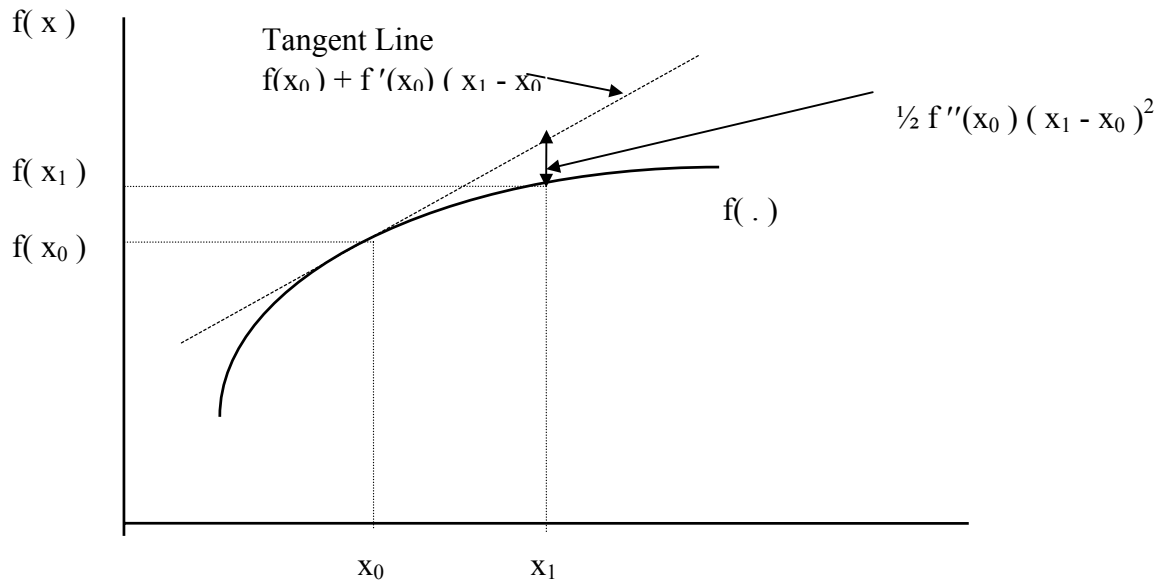
Taylor Series Expansion

A second order Taylor Series expansion of the scalar function $f(x)$ around the point x_0 is:

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + r.$$
 Where r is a remainder term.

$$f(x) \cong \underbrace{f(x_0) + f'(x_0)(x-x_0)}_{\text{Tangent at } x_0} + \underbrace{\frac{1}{2} f''(x_0)(x-x_0)^2}_{\text{2nd order term}}$$

Taylor Series Expansion of $f(x)$ at x_0



Matrix Derivatives

For example, for the **Linear form**:

$$\text{If } L(x) = c'x \quad \text{then} \quad \frac{\partial L(x)}{\partial x} = c',$$

and for the **Quadratic form**:

$$\text{If } Q(x) = x'Ax \quad \text{then} \quad \frac{\partial Q(x)}{\partial x} = 2x'A,$$

In both cases the derivative of a scalar with respect to a column vector is a row vector.

The Gradient Vector ($\nabla f(x)$)

A second convention is that the derivative of a scalar with respect to a column (row) vector is a row (column) vector. Thus, if the scalar y is a differentiable function of the column vector x , the vector of partial derivatives $\frac{\partial y}{\partial x_i}$ is a row vector called the Gradient

vector. For example, if an objective function is a scalar value which is a nonlinear function of n variables.

$$y = f(x) = f(x_1, x_2, \dots, x_n)$$

then the vector of first order partial derivatives, the *gradient vector*, is the row vector

$$\nabla f(x) \equiv \frac{\partial f(x)}{\partial x} = \left[\frac{\partial f(\cdot)}{\partial x_1}, \frac{\partial f(\cdot)}{\partial x_2} \dots \frac{\partial f(\cdot)}{\partial x_n} \right]$$

Inner Products

Because we use inner products a lot we will adopt a new clearer notation. Inner product of two vectors $a'b = \text{scalar}$; will now also be denoted $\langle a, b \rangle = \text{scalar}$.

Thus the familiar objective function $c'x = z$ can be written $\langle c, x \rangle = z$.

Hessian and Jacobian Matrices

The derivative of the gradient vector with respect to the $n \times 1$ column vector x is the $n \times n$ ***Hessian matrix***:

$$\frac{\partial^2 f(x)}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial f(x)}{\partial x} \right) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2}, \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \dots \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1}, \frac{\partial^2 f(x)}{\partial x_2^2} \dots \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1}, \frac{\partial^2 f(x)}{\partial x_n \partial x_2} \dots \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

For example, the Hessian matrix of the quadratic form $x'Ax$ is $2A$.

Jacobian matrix:

Similarly the derivative of the column vector of m functions: $g_i = g_i(x)$, where each function depends on the $n \times 1$ column vector x , with respect to x is the $m \times n$ ***Jacobian matrix***:

$$\frac{\partial g(x)}{\partial x} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \dots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \dots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \dots & \frac{\partial g_m}{\partial x_n} \end{pmatrix}.$$

Taylor Series Expansion of a Vector Function

We can now use gradients, Hessians, and the scalar Taylor series to approximate functions of vectors.

Suppose $f(x)$ is a function of the $n \times 1$ vector x .

Given some vector of initial values x_0 we can expand around this vector to approximate the functional value of some other vector x .

$$f(x) = f(x_0) + \langle \nabla f(x_0), (x-x_0) \rangle + 1/2 (x-x_0)' H_{x_0} (x-x_0) + r_0$$

where $\nabla f(x_0)$ is the Gradient of $f(x)$ at x_0

H_{x_0} is the Hessian of $f(x)$ at x_0

$(x-x_0)$ is an n by 1 vector of the differences between x and x_0 .

r_0 is the remainder term of the expansion that could be reduced by

Definite Quadratic Forms.

Some quadratic forms have the property $x'Ax > 0$ for all x except $x = 0$; some are negative for all x except $x = 0$.

Positive Definite Quadratic Form: The quadratic form $x'Ax$ is said to be positive definite if it is positive (>0) for every x except $x = 0$.

Positive Semidefinite Quadratic Form: The quadratic form $x'Ax$ is said to be positive semidefinite if it is non-negative (≥ 0) for every x , and there exist points for which $x'Ax = 0$ and $x \neq 0$

Negative definite and semidefinite forms are defined by interchanging the words "negative" and "positive" in the above definitions. If $x'Ax$ is positive definite (semidefinite), then $x'(-A)x$ is negative definite (semidefinite).

An Introduction to Nonlinear Optimization

Some Non Linear Programming (NLP) Definitions

The Standard N.L.P. Problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{Subject to} & x \in \Omega \end{array} \quad \text{where } x = nx1 \text{ vector}$$

and where Ω denotes the feasible solution set. $\Omega \in \mathbb{R}^n$ or a subset of $\mathbb{R}^n$.

Note: The objective function or the solution set are no longer defined by linear functions.

Local Minima

A point $x^* \in \Omega$ is a local minimum of $f(x)$ on Ω if there is a small distance ε such that
 $f(x) > f(x^*)$ for all $x \in \Omega$ within the distance ε of x^* .

Verbally "The objective function increases in all directions, therefore we are at the optimum point for a minimization problem."

Global Minima

A point $x^* \in \Omega$ is a global minimum of $f(x)$ on Ω if
 $f(x) > f(x^*) \quad \forall x \in \Omega$

The aim is to set criteria for a computer program to perform a systematic search over a mathematical surface. As in finding your way to a location, good directions will give you a sequence to follow. **In each sequence or step you need to know the direction "d" to proceed, and how far " α " to go in that direction.**

Feasible Directions

Along any given single direction "d" the objective function $f(x)$ is a function of the distance moved in a direction. Note any direction in $\mathbb{R}^n$ is an n dimensional vector.

Definition

For $x \in \Omega$, d is a feasible direction at x_0 if there exists a scalar $\alpha > 0$,

such that $(x_0 + \alpha d) \in \Omega$ for all α , $0 \leq \alpha \leq \bar{\alpha}$.

i.e. A particle can move in direction "d" for a distance $\alpha \leq \bar{\alpha}$ without leaving the feasible set Ω .

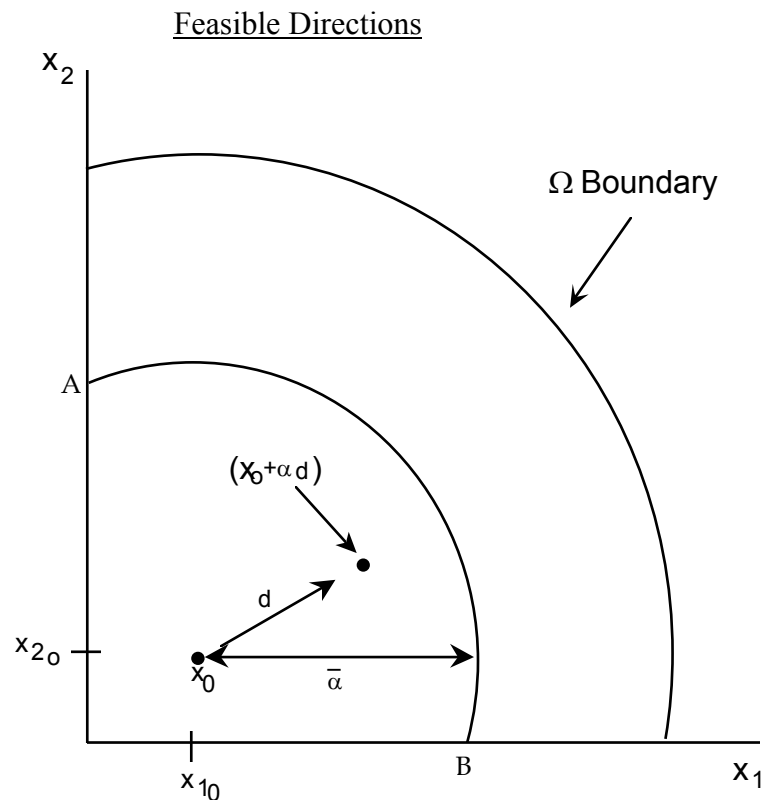
- Note
1. A direction in “n space” (R^n) is an n-dimensional vector “d”.
 2. Along any given single direction d, the objective function $f(x)$ is a function of the distance moved in that direction.

An Example in “Two Space” R^2 where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

d = the directional vector can equal $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$, or $\begin{bmatrix} 4 \\ 2 \end{bmatrix}$, or $\begin{bmatrix} 34 \\ 17 \end{bmatrix}$

Note: The vector d gives a direction, but not the distance .

Verbally "A direction depends on the ratio of values in the vector, not the values themselves."



The feasible directions depend on

- (a) $\bar{\alpha}$ = the stepsize limit
- (b) Where starting point x_0 is situated in Ω .
- (c) In this example, the arc from point A to point B describes the set of feasible directions at x_0 for stepsize $\bar{\alpha}$.

Note: The feasible directions "d" at a point x_0 are a function of the stepsize limit $\bar{\alpha}$ and vice versa.

Nonlinear First Order Conditions

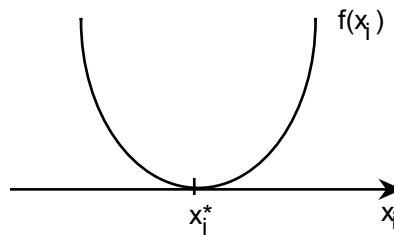
Local Minimum Point–Constrained Problem

Given the objective function $f(x)$ and the constraint set Ω . If x^* is a local minimum point of $f(x)$ over Ω , then for any feasible direction from x^* , the necessary condition is:

For all feasible directions $d \in R^n \cap \Omega$

$$(1) \quad \langle \nabla f(x^*), d \rangle \geq 0$$

Minimizing a Scalar Function



Equation (1) can be explained from the figure as follows.

Moving left from x^* , $d < 0$ and $\nabla f(x) = \frac{\partial f(x)}{\partial x} < 0$ so $\nabla f(x) \cdot d > 0$.

Moving right from x^* , $d > 0$ and $\nabla f(x) = \frac{\partial f(x)}{\partial x} > 0$ so $\nabla f(x) \cdot d > 0$.

This works since d is defined as the direction you're moving from the point x^* , that is, each element is a distance $(x_i - x_i^*)$ as in the Taylor series expansion.

Proof of First Order Condition

Let x^* = a local minimum.

Pick another point $x(\alpha)$ at an arbitrary distance α in direction d ($\alpha > 0$).

$$x(\alpha) = x^* + \alpha d. \quad \therefore \text{The new objective value is } f(x^* + \alpha d).$$

Apply Taylor expansion, truncated a first order to $f(x(\alpha))$ around $f(x^*)$.

$$(2) \quad f(x(\alpha)) = f(x^* + \alpha d) \approx f(x^*) + \langle \nabla f(x^*), (x(\alpha) - x^*) \rangle \approx f(x^*) + \langle \nabla f(x^*), \alpha d \rangle$$

Note that the second order terms and the remainder term in the T.S. expansion have been truncated, making this an approximation.

If x^* is a minimum point, by definition:

$$(3) \quad f(x^*) - f(x(\alpha)) \leq 0$$

and substituting the expansion for $f(x(\alpha))$ defined in (2) into (3), we obtain:

$$(4) \quad f(x^*) - f(x^*) - \langle \nabla f(x^*), \alpha d \rangle \leq 0$$

$$(5) \quad \therefore -\langle \nabla f(x^*), \alpha d \rangle \leq 0$$

factoring out the scalar α and multiplying by -1 we get:

$$(6) \quad \alpha \langle \nabla f(x^*), d \rangle \geq 0 \quad \text{but since } \alpha > 0,$$

$$(7) \quad x^* \text{ being a minimum} \Rightarrow \langle \nabla f(x^*), d \rangle \geq 0.$$

The Unconstrained Problem

If the problem is unconstrained, this implies that x^* is an **interior point** and therefore for some small enough $\alpha > 0$ all directions d are feasible.

Point. For unconstrained problems the feasible direction vector d can have any sign or direction, thus,

The first order condition $\langle \nabla f(x^*), d \rangle \geq 0$ for all d contained in $\mathbb{R}^n$

which implies that $\Rightarrow \nabla f(x^*) = 0$ because:

$$\langle \nabla f(x^*), d \rangle = [f'_1, f'_2, \dots, f'_n] \begin{bmatrix} \pm \text{anything} \\ \pm \text{anything} \\ \vdots \\ \pm \text{anything} \end{bmatrix} \geq 0 \text{ for all possible directions } d$$

This means $[f'_1, f'_2, \dots, f'_n]$ must equal $[0, 0 \dots 0]$

Example A. Unconstrained Problem

$$\text{Min } f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2$$

There are no constraints therefore the feasible region is the whole of “two space” on the real line in symbolic terms: $\Omega = \mathbb{R}^2$

$$\text{F.O.C} \quad \nabla f(\bullet)' = 0 \quad \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} = \text{set} \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solving the gradient equations yields:

$$\begin{aligned} \therefore 2x_1 - x_2 &= 0 & \therefore x_1 &= 1/2 x_2 \\ -x_1 + 2x_2 &= 3 & \text{and } -x_1 + 4x_1 &= 3 \\ \therefore \text{solving for } x_1 \text{ and } x_2 \text{ we get } x_1 &= 1 & x_2 &= 2 \end{aligned}$$

Try substituting this into $f(x_1, x_2)$ and check other values.

Example B. Constrained Optimization

$$\begin{aligned} \text{Min} \quad f(x_1, x_2) &= x_1^2 - x_1 + x_2 + x_1 x_2 \\ \text{subject to} \quad x_1 &\geq 0 \quad \text{and} \quad x_2 \geq 0 \end{aligned}$$

$$\text{Problem B has a global minimum at} \quad \begin{aligned} x_1 &= [1/2] \\ x_2 &= [0] \end{aligned}$$

Do the constrained first order conditions hold at this point?

$$\text{check FOC} \quad \nabla f(x^*) = [2x_1 - 1 + x_2, x_1 + 1]$$

Substituting the numerical $x_1 = 0.5$ and $x_2 = 0$ for x_1 and x_2 in the equation yields the gradient at the x values of $[0.5, 0]$ of $[0, 3/2]$

Now we pick a small stepsize α , say $1/10$

$\therefore$ at x^* the feasible directions given the constraints and the initial values are:

$$d = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} \text{anywhere} \\ \text{positive values only} \end{bmatrix}$$

The condition $\langle \nabla f(x^*), d \rangle > 0$ Implies that the product $[0, 3/2] \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} > 0$

or $3/2 * \Delta x_2 > 0$

While Δx_1 can be positive or negative, Δx_2 can only have positive values, therefore the constrained First Order Conditions in (6) hold.

Steepest Descent Algorithms

Steepest Descent Direction

Proposition. The gradient vector of a function indicates the direction of movement which results in the greatest change in the function value:

Proof. See directional derivative Greenberg “ Foundations of Applied Mathematics “, p. 156-157

Example

Quadratic function $f(x) = a'x + x'Bx$

$$\text{where } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad a = \begin{bmatrix} 10 \\ 8 \end{bmatrix} \quad B = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$$

$$f(x_1, x_2) = 10x_1 + 8x_2 - 2x_1^2 - x_2^2$$

$$\text{Given the Initial Values} \quad x_0 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad f(x_0) = 23$$

$$\text{The gradient at } x_0 \text{ is: } \nabla f(x_0) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2} \right] = [10 - 4x_1, 8 - 2x_2] = [6, 2]$$

$\therefore$ the gradient at x_0 indicates a direction which is a ratio of 3:1 in x_1, x_2 space.

Numerical Case. Suppose we have an additional 2 units to allocate between x_1 and x_2 at x_0 .

Strategy A. Put All Two Units on the Most Profitable Activity,
i.e., x_1 that has the largest marginal product.

$$\therefore \tilde{x}_1 = 1 + 2 = 3 \qquad \tilde{x} = 3$$

$$f(\tilde{x}) = 30 + 24 - 18 - 9 = 27$$

Strategy B. Even Split between x_1 and x_2
ie, One unit is added to each

$$\therefore \bar{x}_1 = 2 \qquad \bar{x}_2 = 4.$$

$$\therefore f(\bar{x}) = 20 + 32 - 8 - 16 = 28$$

This result is an improvement over strategy A.

Strategy C. Use the Gradient Ratio to set the Allocation

The gradient at x_0 is $\nabla f(x_0) = [6, 2]$.

The ratio is 3:1, thus of the two extra units, $\left\{ \begin{array}{l} 1.5 \text{ goes to } x_1 \\ .5 \text{ goes to } x_2 \end{array} \right.$

The new values of x are:

$$\therefore \hat{x}_1 = x_1 + 1.5 = 2.5 \qquad \hat{x}_2 = x_2 + .5 = 3.5 \text{ which yields an objective value}$$

$$f(\hat{x}) = 25 + 28 - 12.5 - 12.25 = 28.25 \quad \text{An additional improvement over strategy B.}$$

An Outline of the Gradient Algorithm

Step 1. Pick an initial starting point x_0 .

Step 2. Check if $\nabla f(x_0) = 0$ if so, stop. We are at a "critical point"

Step 3. If $\nabla f(x_0) \neq 0$ i.e., < 0 for a minimization problem we move to another point

$$x_1 = x_0 + \alpha_1 d_1$$

Where $d_1 = \text{direction} = \nabla f(x_0)$
and $\alpha_1 = \text{step size}$

The objective function improvement condition $f(x_0 + \alpha_1 d_1) < f(x_0)$ holds if and only if $\langle \nabla f(x_0), d \rangle < 0$ for a minimization problem.

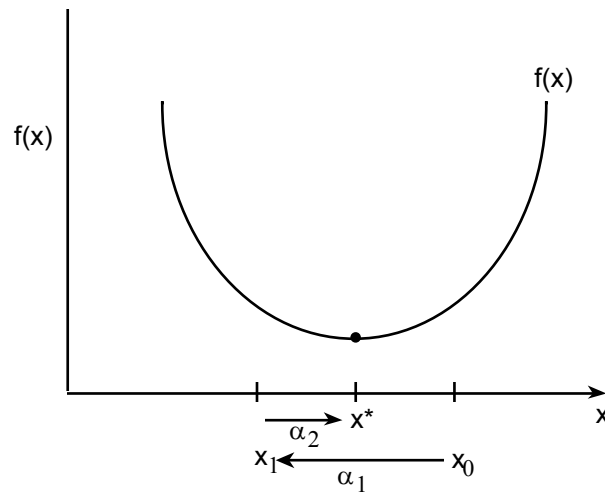
Note

(i) That this improvement condition is analogous to the negative r_j criterion for entering activities in a minimizing LP problem

(ii) The objective function improvement condition has the opposite sign to the first order optimality condition in equation (7) on page 77. This follows since the optimum is defined as the point where no improvement of the objective function is possible.

Step 4. Return to Step 2.

A two dimensional example



(1) Objective – Select x to minimize $f(x)$

(2) Start at x_0 . $\nabla f(x_0) = \text{slope} > 0$, so choose $d < 0$ to satisfy the objective function improvement condition $\langle \nabla f(x_0), d \rangle < 0$.

(3) With stepsize α_1 we arrive at x_1 . At x_1 $\nabla f(x_1) < 0$, so we choose $d > 0$ to satisfy $\langle \nabla f(x_1), d \rangle < 0$.

(4) With stepwise α_2 we arrive at x^* . Note, in this example I faked the selection of α_1 and α_2 .

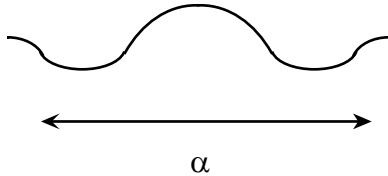
(5) At x^* , $\nabla f(x^*) = 0$ and $\langle \nabla f(x^*), d \rangle \geq 0$. Therefore we are at the minimum.

Practical Problems with Gradient Algorithms

1. **Starting points** must be feasible.
 - Hint: linearize it crudely, find points in the feasible set, use them as starting point.
 - GAMS gives initial conditions that are feasible.
2. **Step size**
 - The stepsize must not take you out of feasible set at each iteration

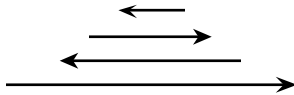


Here α is small and will require many iterations but it will work



Here α is so big you miss the optimum

A self-adjusting step size gets smaller as it moves in towards the optimum.



GAMS/MINOS defaults to this system.

3. **Scaling**

Scaling the values of the data to balance the Hessian matrix of the objective function is a very important operation for all nonlinear solver routines. The essential aspect is to scale the data values and the corresponding coefficients so that the eigenvalues of the Hessian are within three orders of magnitude of each other. Technically, the "condition value" for any matrix is the ratio of the largest and smallest eigenvalue.

An example is the best way of showing this.

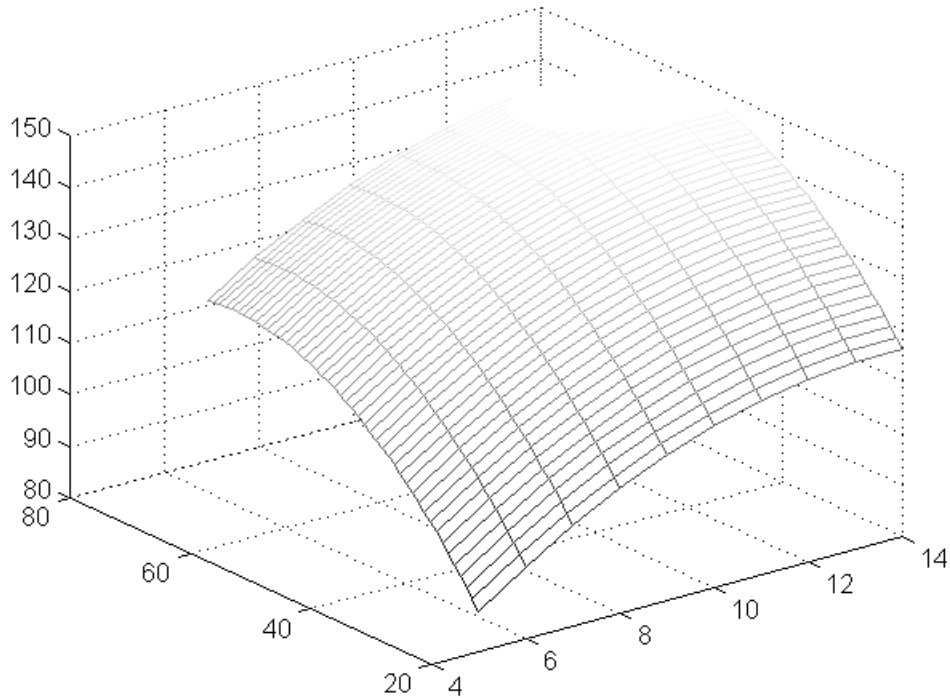
A very simple unconstrained quadratic problem can be defined as:

$$Z = [\alpha_1 \quad \alpha_2] \begin{bmatrix} X \\ Y \end{bmatrix} - [X \quad Y] \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}$$

In case 1 **Well Scaled:** $\alpha = [10.585, 2.717]$ and $\Gamma = \begin{bmatrix} 0.3786 & 0.00578 \\ 0.00578 & 0.02193 \end{bmatrix}$

The eigenvalues for this matrix are 0.3787, 0.0218, the condition number is therefore 17.37.

The surface to be searched by the algorithm appears as follows:

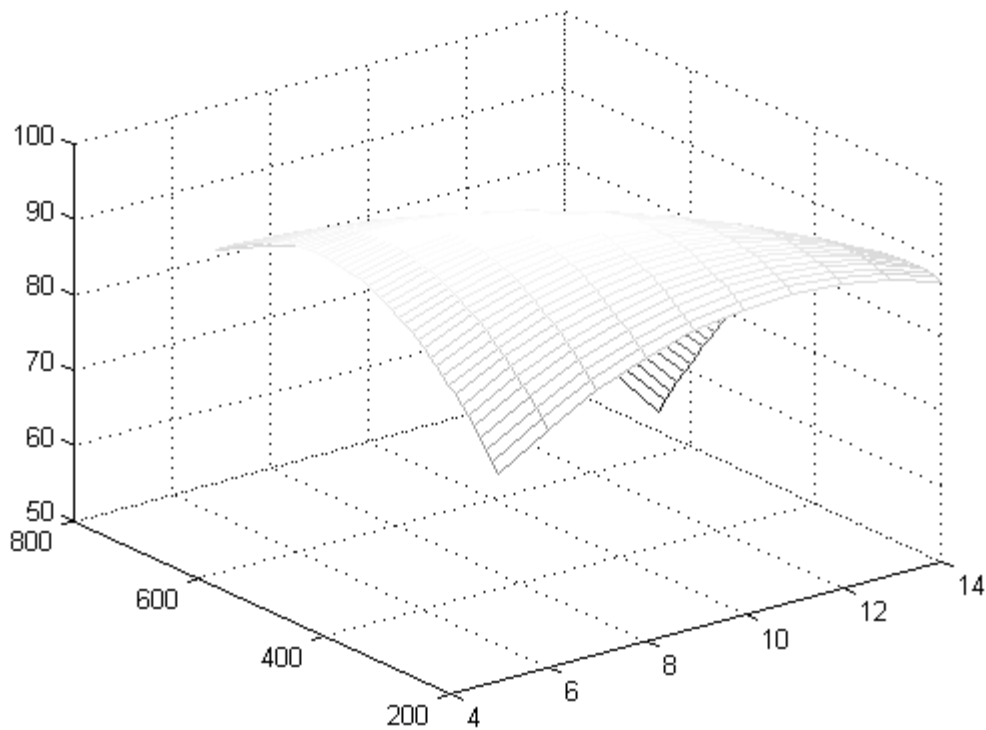


Note that the surface gradients with respect to X and Y are quite similar and the calculation of a gradient will have a similar rounding error in each direction.

Case 2 **Badly Scaled:** $\alpha = [10.585, 0.2717]$ and $\Gamma = \begin{bmatrix} 0.3786 & 0.00578 \\ 0.00578 & 0.000219 \end{bmatrix}$

The eigenvalues for this matrix are 0.3787, 0.0001, the condition number is therefore 3787. Note that a scaling of 10 on variable Y has increased the condition number 218 times.

The surface to be searched by the algorithm appears is now:



Note that the gradients are very different and the Y axis values are multiplied by 10. The same rounding errors and step size are applied in both directions, but have very different effects on the change in the objective Z value. Thus the gradient will try and converge in one dimension but not the other.

This "poorly scaled" condition creates a "hill" on the objective function surface that is very narrow, and the algorithm may "fall off" the surface. Given that the rounding error

in calculating the numerical gradients is the same for the very large and small values the proportional error will be magnified by large differences in scaling.

Reduced Gradients

For constrained nonlinear optimization problems it is very common for the number of nonzero activities in the optimal solution to be greater than the number of binding constraints. Thus, there may be m binding constraints and k ($k \geq m$) activities in the optimum solution, as in the case where we could use the PMP approach. However, all k activities enter the constraint set despite the m dimensional basis. The reduced gradient is the nonlinear analog to the reduced cost in L.P., and incorporates the linkages between activities due to the constraints. The net effect of a marginal change must therefore consider the direct gradient and the effect on other gradients, due to the linkage of the constraints.

Given the constrained nonlinear problem

$$\text{Min } f(x)$$

$$\text{Subject to } Ax \geq b \quad x \geq 0 \text{ and } qx=1$$

At any point, the $qx=1$ vector x can be partitioned into three sets.

- x_B a $m \times 1$ set of basis (dependent) variables
- x_N a $k \times 1$ set of independent variables
- x_0 a $(q-m-k) \times 1$ set of zero valued variables.

Likewise, the matrix A can be partitioned into three matrices B , $m \times m$ basis matrix, N an $m \times k$ matrix of technical coefficients for the independent variables and D , an $m \times (q-k-m)$ matrix for the zero valued variables. Without loss of generality we can drop the zero valued variables (x_0) and the nonbinding constraints from the problem. The problem becomes

$$\text{Min } f(x_B, x_N)$$

$$\text{Subject to } Bx_B + Nx_N = b$$

Using the constraint, the dependent variables can be written as a function of the independent variables.

$$x_B = B^{-1}b - B^{-1}Nx_N$$

Substituting this expression back into the objective function produces two useful characteristics.

- (1) The effect of the binding constraints are incorporated in the objective function.
- (2) The whole problem is expressed as a function of only the independent variables (x_N).

$$\text{Min } f(B^{-1}b - B^{-1}Nx_N, x_N)$$

The resulting "Reduced Gradient" for this problem is:

$$r_{x_N} = \nabla f_{x_N}(\cdot) - \nabla f_{x_B}(\cdot) B^{-1}N$$

The reduced gradient captures the net effect of a marginal change in an independent variable.

Necessary Condition

It can be shown (Luenberger) that a necessary condition for a linearly constrained nonlinear optimization is that all the reduced gradients are zero.

Note. The GAMS/MINOS algorithm uses reduced gradients for this purpose and prints the "rg" value on the right hand side of the activities for nonlinear problems.

Newton's Method

Newton's Method uses the Hessian as well as the gradient to search over the set of all the critical stationary points, i.e., where $\nabla f(x) = 0$. Since we now consider a sequence of algorithm steps, we use the more general notation of $k, k+1, k+2$, etc.

Derivation:

1. Starting at some point x_k , we wish to move to a new point x_{k+1} , which has the property that it is a critical point

$$\nabla f(x_{k+1})' = 0$$

In this method we start with the **gradient** and expand around it.

2. Using a Taylor Series expansion of the gradient around x_k (first-order)

$$\nabla f(x_{k+1})' \approx \nabla f(x_k)' + H_{x_k} (x_{k+1} - x_k) \stackrel{\text{set}}{=} 0$$

where H_{x_k} is the Hessian of $f(x)$ at x_k .

From step (2) we see that the stationary point condition $\nabla f(x_{k+1}) = 0$ implies that:

$$3. \quad H_{x_k} (x_{k+1} - x_k) = -\nabla f(x_k)'$$

Multiply (3) by $H_{x_k}^{-1}$ (assuming that the Hessian is nonsingular and invertible) and move x_k to the right hand side to yield:

$$x_{k+1} - x_k = -H_{x_k}^{-1} \nabla f(x_k)'$$

$$\therefore x_{k+1} = x_k - H_{x_k}^{-1} \nabla f(x_k)'$$

Note: With a quadratic objective function, a stationary point x_{k+1} is reached in one step! An example of Newton's (1642 - 1727) mind at work, which is still topical after three hundred years.

Beautiful!

Example: Newton's Method applied to a Quadratic Problem

$$f(x) = a'x + 1/2 x'Bx$$

where the hessian matrix B is nonsingular and positive definite at x_0 .

Taking the gradient $\nabla f(x_0)' = (a + Bx_0)'$

$$H_{x_0} = B \quad (\text{Note, this is independent of choice of } x_0)$$

$\therefore$ Apply Newton's Method

$$x_1 = x_0 - B^{-1}(a+Bx_0) = -B^{-1}a \quad (\text{Also independent of choice of } x_0).$$

Check. Is x_1 a stationary point?

Plug in $\nabla f(x_1)' = a + Bx_1$ Substitute the Newton value for x_1 above.

$$= a + B(-B^{-1}a)$$

$$= a - a = 0.$$

∴ We have reached a stationary point in one step.

Note: For nonquadratic functions, the Newton is an approximation to the function and therefore will take several steps. However, if the Hessian is "well conditioned," Newton's method will converge rapidly.

Caveats on Newton's Method

1. The Newton approach finds stationary points, it does not guarantee that they're global or even local optima.
2. The Hessian is often hard to invert.

Criteria for Non-Linear Algorithms

1. Theoretical convergence rate
Newton is faster than the steepest decent method
2. Computational difficulty
Newton requires generating, then inverting the Hessian
3. Numerical stability
Newton requires well-conditioned Hessians

Newton Stepsize: Used when the objective function is not quadratic in form. In this case the Taylor expansion that forms the basis of the Newton derivation is only an approximation of the function.

$$x_{k+1} = x_k - \alpha H_{xk}^{-1} \nabla f(x_k)' \text{ for a minimization problem.}$$

Note: If H_{xk} is an identity matrix, then this is equivalent to the steepest descent method and reduces to:

$$x_{k+1} = x_k - \alpha \nabla f(x_k)'$$

Necessary conditions for movement towards a minimum with the Newton method are clarified by a change to condensed notation.

$$\text{Define the notation: } d_k = -H_{xk}^{-1} \nabla f(x_k)' = \text{direction}$$

$$M_k = H_{xx}^{-1} \text{ for simplicity of algebra}$$

$$g_k = \nabla f(x_k)' \text{ for simplicity of algebra}$$

For minimum: We require the "downhill" condition in the new notation to be,
 $\langle d_k, g_k \rangle < 0$ i.e., $d_k' g_k < 0$

Using the notation defined above this is equivalent to:

$$(-M_k g_k)' g_k < 0$$

or
$$g_k' M_k g_k > 0$$

This is most easily fulfilled if $M_k = H_{xx}^{-1}$ is a positive definite matrix.

Desired Conditions for H_{xx}

1. Non-singular
2. Positive definite
3. Well-conditioned (eigenvalues within 10^3 of each other)

How to make an Ill-conditioned Hessian Well-conditioned:

1. Greenstadt's Method - Similar to ridge regression

2. Scaling - a much better approach that aims to change the units of measurement associated with the Hessian so they are "close" i.e., within (10^3) of each other.

The Gams solver Conopt2 now has an effective scaling routine that updates the scaling factors as the solver progresses. This scaling option should not be confused with the standard Gams scaling option that scales the Gams problem before the solver is called. The Conopt2 scaling system has to be called from an appropriate "Options" file in the Gams program. For an example of a Conopt2 options file see chapter 10.

For a more extensive discussion on scaling see:

Gill, Murray, and M. Wright, "Practical Optimization," Academic Press, pp. 346-354.